

ВОПРОС-ОТВЕТ

М.Р. Ефимова

Статистика

www.infra-m.ru

ВОПРОС – ОТВЕТ
серия основана в 2000 г.

Статистика

УЧЕБНОЕ ПОСОБИЕ

Под ред. проф. М.Р. Ефимовой

2223

Москва
ИНФРА-М
2003

УДК 311(075.8)
ББК 65.051я73
С 78

Статистика: Учеб. пособие/Под ред. проф. М.Р. Ефимовой. — М.:
С 78 ИНФРА-М, 2003. — 336 с. — (Серия «Вопрос-ответ»).

ISBN 5-16-000308-8

В учебном пособии рассмотрен широкий спектр статистических методов: группировки, обобщающие статистические показатели, методы изучения взаимосвязей с помощью корреляционно-регрессионного и индексного методов анализа. Излагаются вопросы использования выборочного метода в связи с расширением сферы его применения в статистической практике, особое внимание уделено методам анализа временных рядов и прогнозированию.

В ряде глав использованы реальные статистические данные, характеризующие социально-экономическую ситуацию в России.

Книга предназначена для студентов экономических вузов и финансистов, особенно заочной и дистанционной форм обучения, а также для самообразования предпринимателей и работников финансовых и банковских структур.

ББК 65.051я73

ISBN 5-16-000308-8

© Коллектив авторов, 2000

ПРЕДИСЛОВИЕ

Предлагаемое читателю учебное пособие «Статистика» написано в соответствии с программой курса «Общая теория статистики», являющегося одной из основных дисциплин в системе экономического образования. Мы сочли целесообразным, учитывая объективный рост числа обучающихся по вечерней и заочной формам обучения, а также расширение сфер применения дистанционного образования, построить изложение вопросов теории статистики в более дробном виде, чтобы помочь читателям в усвоении материала. Для облегчения понимания предмета вопросы статистической методологии излагаются на страницах учебного пособия в максимально доступной форме в виде вопросов и ответов. Большое количество примеров, различных форм графического представления данных также должны способствовать усвоению курса.

Книга может быть полезной и для самообразования предпринимателей, экономистов и управленцев всех уровней, поскольку в условиях рыночной экономики резко возрастает спрос на услуги статистики. В современных условиях органы государственного и муниципального управления постепенно приходят к осознанию необходимости опоры на статистическую информацию для повышения качества управленческих решений. Субъектам малого и среднего бизнеса необходимы аналитические материалы о состоянии товарных рынков, результатах маркетинговых исследований и т.д.

Знание основных статистических показателей — средних величин, показателей вариации, индексов, коэффициентов корреляции и регрессии и др. — является необходимым условием анализа социально-экономических процессов, в том числе при изучении таких новых для российской экономики явлений, как характеристика состояния и тенденций фондовых рынков, анализ и измерение финансовых рисков и т.д.

Учебное пособие знакомит с общими правилами обработки статистических данных, основными статистическими методами анализа и границами их применения.

Книга состоит из 9 глав, в которых рассмотрены задачи статистического изучения действительности, важнейшие функции государственной статистики, проблемы получения статистической информации, значение метода группировок как основы системного анализа собранных данных, а также роль обобщающих статисти-

ческих показателей, изучение их изменений во времени, прогнозирование с помощью трендовых и регрессионных моделей, исследование взаимосвязей с помощью индексного метода.

В написании учебного пособия принимали участие:

доц. Беликова Г.С. (параграфы 1.4, 1.5, 2.3, 2.5); доц. Купцова Е.В. (параграфы 1.6, 2.1, 2.2); доц. Саблина Е.А. (параграфы 2.6, 3.6, 3.7). Параграфы 4.4. — 4.9. написаны проф. Ефимовой М.Р., параграф 2.4. написан совместно доц. Купцовой Е.В. и Саблиной Е.А. Все остальные параграфы написаны д.э.н., проф. Ефимовой М.Р.

Глава 1
ПРЕДМЕТ И ЗАДАЧИ СТАТИСТИКИ

1.1. Как развивалась статистика и каковы ее задачи?

Возникновение статистической науки относят к середине XVIII в., хотя практические операции по сбору данных о населении, его составе, имущественном положении и других сведений были известны задолго до возникновения статистической науки. Потребности в статистических данных исторически возникают с развитием государства. С помощью системы разрабатываемых показателей статистика всегда стремилась получить характеристику различных сторон общественной жизни, включая область производства, политических отношений, культуры, быта и т.п.

Термин «статистика» произошел от латинского слова *status* («статус»), что означает состояние, положение вещей. Первоначально статистика давала словесное описание «достопримечательностей» государства и лишь с XIX в. статистические сведения стали сообщать в количественной форме. Сейчас слово «статистика» имеет несколько значений. Прежде всего, оно относится к обозначению собранных статистических данных о различных явлениях общественной жизни. Под статистикой понимают также отрасль практической деятельности, задачами которой являются сбор, обработка и анализ статистических данных. И, наконец, под статистикой понимают специальную научную дисциплину, занимающуюся разработкой теоретических положений и методов, используемых статистической практикой.

Возникновение статистической науки связано с именами английского экономиста В. Петти (1623—1687) и Д. Граунта (1629—1674), статистические и демографические идеи которых были развиты их последователям — немецким пастором И. Зюсмильхом (1707—1767) и виднейшим бельгийским ученым XIX в. А. Кетле (1796—1874).

Работы Адольфа Кетле показали значение статистики в познании закономерностей общественной жизни, указав, что эти закономерности четко проявляются лишь в массе явлений, т.е. при изучении данных по большому числу случаев. В трудах А. Кетле дано общее определение предмета статистики и сформулированы задачи статистического познания. Он активно участвовал в разработке пра-

вил проведения переписей населения и рекомендовал десятилетнюю периодичность их проведения.

Учение А.Кетле о статистической закономерности оказало значительное влияние на развитие статистической методологии в работах немецкого статистика и экономиста В. Лексиса (1837—1914), обосновавшего теорию устойчивости и известное правило сложения дисперсий. В. Лексис был сторонником применения математических методов в экономике. Это направление развития статистической науки нашло отражение в работах Ф. Гамильтона (1822—1911), К. Пирсона (1857—1936), Р. Фишера (1890—1962), В. Госсета (1876—1937) и других западных ученых. Их именами назван целый ряд показателей и критериев, с которыми читателю предстоит познакомиться в следующих главах.

В российской статистической школе развитие предмета и метода статистической науки, понимания ее роли в процессе познания общественных явлений шли по трудному и порой противоречивому пути, и до сих пор эти проблемы являются дискуссионными. В основных трудах русских статистиков XIX и начала XX в. статистика рассматривалась как общественная наука.

В работах К.Ф. Германа (1767—1838) обосновывалась необходимость выделения особой науки — теории статистики, функции которой, по его мнению, должны были состоять не только в регистрации фактов, но и в их обобщении и анализе.

В истории развития статистической науки XIX в. важная роль принадлежит работам Д.П. Журавского (1810—1856), в которых он называет статистику наукой «категорического вычисления», подчеркивая значение метода группировок в статистических исследованиях.

Широкую известность приобрели во второй половине XIX в. труды профессора Петербургского университета Ю.Э. Янсона (1835—1893), который был не только видным теоретиком в области статистики (его перу принадлежал и фундаментальный учебник «Теория статистики»), но и руководителем практических статистических работ — в 1881 и 1890 гг. он организовал переписи населения Петербурга.

Конец XIX в. и начало XX в. — наиболее яркие этапы развития российской статистики. С одной стороны, в это время практические потребности пореформенной России явились мощным катализатором развития статистической науки, с другой стороны, в это время идеи западных ученых активно анализируются и обсуждаются российскими статистиками, и, наконец, что особенно важно, эти годы ознаменовались крупными успехами Петербургской школы.

П.Л. Чебышев (1821—1894), а также его ученики А.М. Ляпунов (1875—1918) и А.А. Марков (1856—1922) создали русскую школу

теории вероятностей, достижения которой стали использоваться первоначально в демографии, а затем и в различных областях социально-экономической статистики. Крупнейшим теоретиком статистики начала XX в. был профессор Петербургского университета А.А. Чупров (1874—1926). Его работы, способствовавшие развитию математической статистики, получили признание не только в России, но и за рубежом. Нам представляется важным и для настоящего времени его определение предмета статистического познания. А.А. Чупров относил статистику, с одной стороны, к номографическим наукам, которые изучают общие закономерности развития массовых явлений, а с другой стороны, подчеркивал ее принадлежность к идеографическим наукам, объясняющим закономерности конкретных явлений в определенных условиях места и времени.

Ученица А.А. Чупрова профессор Н.М. Виноградова, разделявшая и развивавшая его взгляд на предмет статистической науки, методы статистического исследования взаимосвязей, более 30 лет заведовала кафедрой статистики в Московском инженерно-экономическом институте (теперь Государственный университет управления).

Вопросы совершенствования методов статистического исследования для принятия управленческих решений на уровне предприятий, различных отраслей и экономики в целом нашли отражение в работах видных российских ученых XX в.: академика В.С. Немчинова, чл.-корр. Академии наук С.Г. Струмилина, Б.С. Ястремского, Д.В. Савинского, П.П. Маслова, Н.К. Дружинина, Г.И. Бакланова, В.Е. Адамова и многих других.

Переход России к рыночным отношениям, смена общественно-политического устройства общества привели к использованию новых методов управления экономикой, и, как следствие, к переориентации деятельности статистических органов и изменению задач статистики. В современных условиях статистика должна обеспечивать органы государственного управления всех уровней, международные организации, коммерческие структуры и население объективной и полной информацией по вопросам социально-экономического развития Российской Федерации, ее регионов, отраслей и секторов экономики. Статистическая информация используется на макроуровне для разработки экономической политики в стране, формирования проектов бюджета Российской Федерации и ее субъектов, получения прогнозов социально-экономического развития, анализа реализации федеральных законов и международных обязательств России.

В связи с переходом к рыночной экономике важной задачей органов негосударственной статистики стало введение в статисти-

ческую практику показателей и классификаций, предназначенных для анализа процессов в рыночной экономике. В рамках этой задачи осуществляется переход к системе национальных счетов (СНС), используемых для характеристики общей картины состояния и развития экономики на макроуровне, установления взаимосвязей между наиболее важными макроэкономическими показателями, к которым относятся валовой внутренний продукт, конечное потребление, валовое накопление, национальное богатство и др.

В современных условиях важное значение приобретает такая задача статистики, как информирование населения, общественно-политических организаций, научно-исследовательских и образовательных учреждений о процессах, происходящих в экономическом и социальном развитии общества.

Сегодня не только предприниматели, научные работники, менеджеры предприятий, но и каждый гражданин России должен владеть основами статистических знаний. Без этого сложно разобраться в программах различных партий и блоков, документах правительства по вопросам социально-экономического развития, сообщениях электронных средств массовой информации и периодической печати. Владение основами статистики позволяет ориентироваться в подчас противоречивых публикациях, критически осмысливать приводимые в различных изданиях статистические данные, грамотно интерпретировать данные органов государственной статистики.

Особенно внимательно приходится относится к тем новым процессам, отражение которых лишь с недавнего времени стало задачей российской статистики. Так, переход к рыночной системе хозяйствования привел к появлению рынка труда и изменению политики государства в области занятости. В этой связи возникла необходимость разработки системы статистических показателей, оценивающих состояние рынка труда на региональном и федеральном уровнях; измерения уровня безработицы; анализа факторов, влияющих на занятость населения в стране; выявления основных тенденций в изменении численности занятого населения по регионам России и отраслям экономики и т.д.

Такая информация для органов власти и государственного управления важна с точки зрения разработки мер по снижению социальной напряженности в обществе, необходимости предусмотреть в достаточном объеме финансовые ресурсы на выплату пособий по безработице, переобучение безработных, содержание служб занятости и прочие потребности.

Став членом целого ряда международных экономических организаций, Россия приняла на себя обязательства по обмену информацией с Организацией Объединенных Наций, Международным

валютным фондом, Всемирным банком, Европейским банком реконструкции и развития, Международной организацией труда и др.

В связи с научной организацией управления предприятиями и организациями возрастает роль статистики при изучении и анализе работы предприятий и решении многих важных задач организации производства, контроля качества продукции, разработки норм и нормативов и т.д.

Очень емким является определение роли статистики на предприятии, данное французским ученым Р. Дюма: «Статистика необходима для того, чтобы знать, для того, чтобы предвидеть, для того, чтобы действовать, для того, чтобы проверять»[1].

Каждое предприятие разрабатывает текущие и перспективные планы, основу которых в условиях рыночной экономики составляют прогнозы. Так, начальный этап планирования связан с оценкой возможности предприятия по выпуску продукции, когда учитываются не только уже сложившаяся динамика объема продукции, но также ее конкурентоспособность, состояние рынка, динамика цен и др., т.е. факторы, которые могут повлиять на изменение планируемого показателя. К этим факторам относятся также эффективность использования рабочей силы и средств труда, качество управления, в том числе компетентность и квалификация кадров, совершенство применяемой техники и технологии и многое другое.

Многогранная характеристика различных сторон деятельности предприятий, отраслей и национальной экономики находит отражение в системах статистических показателей, которые можно назвать языком статистики. Задача статистики состоит в том, чтобы разрабатывать методологию построения этих показателей, совершенствовать и развивать методы их анализа, объяснять, как правильно интерпретировать полученные данные и давать оценку того, насколько ценна для практического использования полученная информация.

1.2. На какие вопросы должен отвечать статистический показатель?

Как было показано, одна из главных функций статистики состоит в обеспечении информацией, которая может использоваться в различных ситуациях. Характерной особенностью статистики является числовое освещение действительности. Каждый статистический

[1] Дюма Р. Статистика предприятия. М., 1958. С. 390.

показатель характеризует определенный объект в конкретных условиях времени и места, т.е. отвечает на вопросы: что? где? когда?

Так, в январе—сентябре 1999 г. (когда?) численность населения с денежными доходами ниже величины прожиточного минимума (что?) в Российской Федерации (где?) составила 49,9 млн человек. Что означает этот показатель, много это или мало, насколько это может быть важно для оценки степени социальной напряженности в обществе? Интерпретация этого показателя вне связи с другими взаимосвязанными показателями весьма затруднена для лиц, не являющихся специалистами в данной области. Но если мы дополним сказанное данными за январь—сентябрь 1998 г., когда этот показатель был равен 33,5 млн человек, становится очевидной тенденция достаточно быстрого роста (почти 49%) численности этой категории населения в течение только одного года. Можно сопоставить данные о численности населения с денежными доходами ниже величины прожиточного минимума с общей численностью населения. Тогда в январе—сентябре 1999 г. доля такого населения составила 34,1% от общей численности населения, тогда как за тот же период 1998 г. она была равна 22,8%, что также свидетельствует о росте доли населения с доходами ниже прожиточного минимума. Приведенный небольшой пример иллюстрирует значимость и широкое распространение сравнений в статистике. Те или иные статистические данные могут сравниваться с показателями предшествующих периодов (сравнения во времени), с показателями, характеризующими другие единицы совокупности (например, сравнение данных по разным предприятиям отрасли), и, наконец, могут проводиться территориальные сравнения (например, сопоставляются показатели уровня жизни населения разных стран).

Опора на статистические показатели, анализ условий их формирования и тенденций изменения позволяет перейти от субъективных мнений и суждений к объективной оценке происходящих в исследуемой области процессов.

1.3. Как определить предмет и метод статистики?

Статистика— наука, характеризующая количественную сторону качественно определенных массовых явлений в конкретных условиях места и времени.

В этом определении подчеркиваются следующие основные особенности статистической характеристики исследуемых явлений. Прежде всего, количественную сторону явлений нельзя рассматривать в отрыве от их качественной определенности. В основу всякого

статистического исследования происходящих в обществе процессов и явлений должна быть положена социально-экономическая теория данного вопроса. Руководствуясь социально-экономической теорией, статистика обогащает ее знанием фактов, показывает конкретные проявления теоретически сформулированных закономерностей в определенных условиях места и времени.

Кроме того, в определении указывается, что статистические показатели не относятся к отдельному случаю, а всегда представляют собой результат обобщения данных по массе случаев. Такую массу случаев называют статистической совокупностью. Статистическая совокупность представляет собой множество существующих объектов, однородных в определенном отношении. Под однородностью здесь понимается сущность явлений, определяемая законами их развития.

Важной чертой статистической совокупности является наличие вариации, т.е. различий в величине признаков, характеризующих различные свойства отдельных единиц, входящих в исследуемую совокупность. Признаки, характеризующие каждую единицу совокупности, отражают частный случай проявления изучаемой закономерности. На основе всего множества ее проявлений статистика может дать количественную характеристику исследуемой закономерности.

Для статистических совокупностей характерно свойство устойчивости, т.е. их характеристики сохраняют определенное значение в течение достаточно длительного промежутка времени. Так, отмеченное Д. Граунтом еще в XVI в. превышение числа родившихся мальчиков над числом родившихся девочек было подтверждено в более поздних исследованиях, т.е. статистика разных стран свидетельствует о том, что на 100 родившихся девочек обычно приходится 106 родившихся мальчиков.

Большая устойчивость характерна для показателей, относящихся к странам, регионам, отраслям экономики. Показатели, с помощью которых оценивается деятельность предприятий и организаций, характеризуются меньшей устойчивостью, поскольку они складываются в значительной степени под влиянием индивидуальных особенностей соответствующих единиц.

Общая теория статистики формирует общие правила и приемы статистического исследования массовых общественных явлений, которые составляют содержание статистической методологии. Общие правила статистического исследования опираются на законы диалектики, согласно которым изучаемые явления и процессы рассматриваются в развитии, взаимной связи и причинной обусловленности. Как всякий научный метод, статистика основывается на соединении анализа и синтеза.

В определении статистики подчеркивалось своеобразие этой науки, состоящее в том, что количественная форма статистических суждений опирается на знание качественной природы изучаемых явлений и процессов. Из этого своеобразия вытекают и особенности статистического анализа, заключающегося в том, что статистические методы исследования органически сочетаются с методами той науки, предмет которой изучается.

Вместе с тем метод статистики включает ряд специфических приемов статистического исследования. К этим приемам относятся:

- массовое статистическое наблюдение;
- сводка и группировка первичных данных статистического наблюдения;
- определение обобщающих показателей;
- анализ, обобщение и интерпретация статистических показателей.

В процессе статистического наблюдения получают исходные данные, служащие основой для получения объективной информации по изучаемой проблеме. Особое значение первого этапа статистического исследования состоит в том, что лишь достаточно достоверные и полные сведения, полученные в процессе статистического наблюдения, позволяют на последующих этапах обеспечить научно обоснованные выводы о тенденциях и взаимосвязях в развитии изучаемого объекта. Любые новейшие и сложные методы анализа не дадут желаемого результата, если их применение не будет основываться на качественных исходных данных. Требования к качеству исходных данных объясняют необходимость организации статистического наблюдения на научной основе.

Характерным для статистики является не изучение отдельных единиц совокупности и их свойств, а выявление на основе наблюдения большого числа единиц существенных черт и взаимосвязей, а также закономерностей в изменении изучаемых явлений во времени и пространстве.

Одним из важных принципов статистики является изучение не только явления в целом, но и стремление показать их дифференцированно, во всем многообразии типов, выявить особенности разных групп единиц, развивающихся в различных условиях. Так, например, при анализе внешней торговли России в составе внешнеторгового оборота выявляют экспорт и импорт, которые затем изучают в разрезе отдельных товарных групп, характеризуют географическое распределение товаров и т.д.

Результаты статистического наблюдения сообщаются в форме статистических показателей. Показателем называют обобщенную количественную характеристику качественно определенных социально-экономических явлений и процессов применительно к конкретным условиям места и времени. В процессе статистического наблюдения с помощью показателей дается характеристика отдельной единицы совокупности, а в результате группировки получают групповые и общие (сводные) показатели. Показатели, дающие характеристику какой-либо отрасли или территории, называют соответственно отраслевыми и региональными.

Совокупность показателей, всесторонне отражающих тот или иной объект, образует систему показателей. Такие системы рассматриваются в социальной, экономической статистике и в статистике отраслей.

Непосредственно по результатам наблюдения фиксируется абсолютная величина, характеризующая размеры или объемы изучаемых совокупностей. Так, в отчетности предприятий приводится объем выпущенной продукции в натуральном или стоимостном выражении, стоимость основных фондов, численность работников и т.д. Целый ряд показателей, характеризующих деятельность предприятий и организаций, являются производными и определяются расчетным путем. Величина таких показателей определяется в соответствии с методологией их расчета. В качестве примера расчетных показателей можно назвать производительность труда, рентабельность продукции, оборачиваемость активов и т.п. Разработка методологии расчета статистических показателей является одной из функций органов государственной статистики.

На заключительной стадии статистического исследования проводится анализ, в процессе которого исследуются особенности структуры объекта, степень однородности данной группы или совокупности в целом, тенденции развития объекта исследования во времени, степень влияния различных факторов на формирование результатов и т.п. Анализ завершается интерпретацией полученных результатов, формулировкой теоретических и практических выводов и предложений.

Разнообразные виды обобщающих статистических показателей и методы их анализ будут рассмотрены в главах 4—9.

1.4. Каковы основные этапы становления российской государственной статистики?

За два века своей истории российская статистика прошла сложный путь. В сентябре 1802 г. в соответствии с Высочайшим манифестом Императора Александра I была введена письменная отчетность министерств. Так началось организационно-структурное

оформление государственной статистики в России. Этот год принято считать годом рождения российской государственной статистики.

В 1911 г. впервые был создан официальный центр правительственной статистики — Статистическое отделение при Министерстве полиции, которое в последующем было передано в ведение Министерства внутренних дел; сюда поступали отчеты из губерний. Первым руководителем Статистического отделения был академик К.Ф. Герман (1767—1838).

Статистические работы проводились и вновь созданными различными министерствами и ведомствами: Министерством внутренних дел, Главным управлением путей сообщения (сведения о перевозках по внутренним водным путям), Министерством коммерции (данные о внешней торговле) и др. Так появилась ведомственная статистика. Высокая квалификация работников министерств позволяла получать статистические работы лучшего качества, чем полицейская статистическая отчетность губерний.

В 1852 г. Статистическое отделение было преобразовано в Статистический комитет, ставший с 1858 г. Центральным статистическим комитетом Министерства внутренних дел (ЦСК).

В 1864—1875 годах ЦСК возглавлял П.П. Семенов—Тян-Шанский — известный географ, экономист, статистик. Под его руководством в 1870 г. был проведен Первый статистический съезд в России, а в 1972 г. — Международный статистический конгресс, на котором П.П. Семенов сделал доклад о принципах организации переписи населения. Конгресс способствовал взаимному обогащению российской и зарубежной статистики. В этот период были выполнены многочисленные статистические работы: подворные обследования крестьянских хозяйств, военно-конские переписи (для военных нужд), перепись паровых двигателей, школьная перепись (1880 г.), имевшая большое значение в развитии статистики просвещения, работы по организации статистики транспорта, внешней торговли. С 1883 г. стали вести статистику урожаев. Важнейшей работой этого времени стала всеобщая перепись населения Российской империи 1897 г., которой предшествовало 98 городских переписей. Всеобщая перепись населения проводилась под руководством П.П. Семенова и стала воплощением его идеи — введения точного учета населения как основы государственной статистики.

Таким образом, в XIX в. в России складывалась система государственной статистики: руководящим органом был ЦСК, местными органами — губернские статистические комитеты и земские статистические бюро, основная работа которых состояла в получении разнообразных данных о губерниях. Земской статистике принадлежит особое место в истории российской государственной статистики. При земствах, органах местного самоуправления с середины 70-х годов были созданы специальные статистические бюро. Земские статистики собирали и разрабатывали огромный статистический материал, который использовался для глубоких экономических исследований пореформенной России. Эти работы сопровождались и развитием статистической методологии.

Видными земскими статистиками были В.И. Орлов (1848—1865), П.П. Червинский (1849—1931), Ф.А. Щербина (1849—1936), А.П. Щликевич (1849—1906), впервые построивший комбинационную таблицу по сочетанию четырех признаков и применившей экономико-статистические методы для характеристики связи различных крестьянских дворов.

Земская статистика стала прообразом современной муниципальной статистики.

В 90-х годах были созданы фабрично-заводские инспекции, которые вели текущую статистику, разрабатывали данные по статистике труда, в том числе о составе рабочей силы, несчастных случаях, стачках и др.

Стала развиваться промышленная статистика. Под руководством В.Е. Варзара (1851—1940) в 1900, 1908 и 1923 г. были проведены первые переписи промышленности. При разработке материалов переписей впервые была применена классификация отраслей промышленности.

В организационном отношении правительственная статистика, сложившаяся в 60-х годах XIX в., сохранялась до 1917 г.

Следующий этап в развитии государственной статистики России начался после 1917 г.

Качественная ломка политических и производственных отношений после революции 1917 г. привела к созданию принципиально иного аппарата государственной статистики, формировавшегося в соответствии с задачами планово-административной экономики.

В июне 1918 г. состоялся Первый Всероссийский съезд статистиков, и было принято «Положение о государственной статистике», в соответствии с которым было организовано Центральное статистическое управление (ЦСУ) на правах народного комиссариата. На него возлагалась задача централизованного руководства как практической статистикой, так и развитием ее методологии. Возглавил ЦСУ видный деятель российской статистики П.И. Попов (1872—1950). В губерниях, городах, уездах были организованы местные статистические учреждения. Таким образом, создавалась единая централизованная система государственной статистики.

Позднее с созданием СССР были учреждены центральные статистические управления в союзных республиках, краевые, областные, районные управления в административных единицах.

С конца 20-х годов по мере восстановления разрушенной гражданской войной экономики все более возрастала роль статистики как средства управления народным хозяйством, а это требовало и совершенствования организации самой статистики. В 1927 г. было принято новое Положение о ЦСУ, направленное на повышение научного уровня статистической работы, усиление ее координации с работой плановых органов.

Начавшийся период индустриализации стал и периодом интенсивного развития статистики промышленности. ЦСУ был проведен ряд крупных статистических исследований (обследование промышленных предприятий, перепись предприятий легкой промышленности и др.). Самой значительной работой стало составление ЦСУ баланса народного хозяйства (1923—1924 гг.). Вместе с тем в нем не была учтена многоукладность экономики того периода.

Претворение в жизнь идеологии жесткой плановой экономики и принятие первого пятилетнего плана поставило перед государственной статистикой задачу контроля за ходом его выполнения, а также своевременного обеспечения плановых расчетов статистическими данными. Это явилось основной причиной реорганизации статистических органов и в 1930 г. ЦСУ было объединено с Госпланом, что было связано с попытками обеспечить более активную роль статистики в планировании и управлении народным хозяйством.

Объединение органов государственной статистики с органами планирования имело как положительные, так и отрицательные стороны. Улучшилась текущая статистика и текущее обеспечение плановых органов необходимыми данными, произошло сближение плановой и статистической методологии. Был проведен ряд крупных работ, в частности, всесоюзные переписи населения 1926, 1937, 1939 гг. Огромная работа по учету материальных и трудовых ресурсов была выполнена за годы Великой Отечественной войны. В то же время подчинение органов государственной статистики органам Госплана привело к снижению роли статистики в государственном управлении, ее основной функцией стала оценка выполнения установленных планов. Меньше стало уделяться внимания глубоким аналитическим статистическим работам. Статистики не всегда имели возможность объективно оценивать реальные экономические процессы.

В 1948 г. в результате отделения органов государственной статистики от Госплана СССР было образовано Центральное статистическое управление при Совете Министров СССР, а в 1978 г. — Центральное статистическое управление СССР.

Экономические преобразования в стране, начавшиеся после 1985 г., привели к очередной реорганизации статистических органов, и ЦСУ СССР было преобразовано в союзно-республиканский Государственный комитет СССР по статистике (Госкомстат СССР).

В связи с распадом СССР в 1992 г. на базе Госкомстата СССР был образован Госкомстат СНГ, который в настоящее время осуществляет координирующую работу в области методологии и получения статистических данных по странам СНГ. Главным государственным статистическим органом России стал Государственный комитет Российской Федерации по статистике (Госкомстат РФ).

Госкомстат РФ со своими административно-территориальными органами составляет единую систему органов государственной статистки. Эта система образована с учетом административно-территориального деления страны. В автономных республиках, входящих в Российскую Федерацию, имеются республиканские комитеты по статистике, в автономных областях, округах, краях, областях, а также городах Москве в Санкт-Петербурге — соответствующие областные, краевые комитеты по статистике, в районах, городах — управления по статистике — первичные звенья централизованной статистической системы.

1.5. В чем состоят функции органов государственной статистики?

Преобразования в российской экономике, связанные с рыночными реформами, обусловили необходимость адекватного изменения как организации государственной статистики, так и ее функций.

Госкомстат Российской Федерации является федеральным органом исполнительной власти. В своей деятельности он руководствуется Конституцией, законами, распоряжениями и постановлениями Президента и Правительства, а также федеральными статистическими программами, которые формируются с учетом предложений федеральных органов исполнительной и законодательной власти, органов государственной власти субъектов Российской Федерации, научных и других организаций и утверждаются Госкомстатом по согласованию с Правительством.

Методология статистических показателей, формы и методы сбора и обработки статистических данных, установленные Госкомстатом, являются официальными статистическими стандартами Российской Федерации.

Госкомстат выполняет следующие основные функции:

- организует проведение государственных статистических наблюдений по разработанной и утвержденной программе, методологии, формам;
- обеспечивает функционирование Единого государственного регистра предприятий и организаций (ЕГРПО) на основе учета всех хозяйствующих субъектов на территории Российской Федерации с присвоением им идентификационных кодов, исходя из общероссийских классификаторов технико-экономической и социальной информации;
- вырабатывает единые подходы к формированию целевых регистров министерствами и ведомствами;
- проводит единую техническую политику в технологии сбора, разработки и передачи статистической информации, взаимодействует с региональными, отраслевыми, информационно-вычислительными системами;
- обеспечивает сбор, обработку, хранение и защиту статистической информации, соблюдение государственной и коммерческой тайн, необходимую конфиденциальность данных;
- осуществляет международные сопоставления основных социально-экономических показателей Российской Федерации, совместно с Центральным банком России составляет платежный баланс на основе данных, полученных от финансово-кредитных, внешнеэкономических и других органов;
- создает условия для эффективного использования средств вычислительной и организационной техники, внедряет новейшие технологии обработки информации, современные лицензионные программные продукты и системы управления базами данных, аналитические и информационные справочные системы;
- участвует в программах сотрудничества с национальными и международными статистическими и экономическими организациями.

Особую значимость в современных условиях приобретает деятельность административно-территориальных и местных органов государственной статистики в связи с развитием региональной и муниципальной статистики.

В числе основных функций региональных статистических органов в настоящее время можно назвать:

- исследование экономических, социальных, духовных, экологических явлений, обобщение и прогнозирование тенденций их развития;
- обеспечение региональных (муниципальных) органов управления необходимой оперативной статистической информацией;

- формирование необходимой информации для разработки региональных, муниципальных программ и мониторинг процессов их реализации;
- совершенствование системы показателей, адекватно отражающей происходящие в регионе процессы;
- создание и ведение региональных регистров, кадастров, пообъектных баз данных для комплексной экономической, демографической, социологической характеристики производителей.

Таким образом, Госкомстат России вместе с территориальными органами представляет собой централизованную иерархическую систему, выполняющую важнейшую функцию — информационное обеспечение государственного и муниципального управления.

1.6. Каковы отрасли статистической науки?

Чтобы управлять современным обществом, следует отслеживать состояние и взаимодействие друг с другом двух основных сфер — экономики и социальной системы. Несмотря на то что эти две сферы тесно взаимосвязаны и взаимозависимы (так, например, уровень экономического развития страны отражается на уровне жизни населения, а кадровый потенциал во многом определяет ход экономических преобразований), каждая из них имеет конкретный, обособленный предмет исследования и, следовательно, требует адекватной системы статистических показателей.

Кроме того, внутренние процессы, протекающие как в сфере экономики, так и в социальной сфере, различны по содержанию и могут классифицироваться в зависимости от макро- и микроуровня, вида деятельности, отрасли и т.д.

Эти обстоятельства предопределили выделение в составе единой статистической науки отдельных разделов и отраслей. Структура статистической науки представлена на рис.1.1.

Как видно из рис. 1.1, в составе статистической науки выделяются два раздела: общая теория статистики и социально-экономическая статистика. Социально-экономическая статистика, в свою очередь, подразделяется на экономическую статистику и ее отрасли, социальную статистику и ее отрасли.

Методологические принципы и методы, применяющиеся при изучении различных сторон общественной жизни, излагаются в разделе «Общая теория статистики». Этот раздел статистической науки посвящен разработке общих методов и правил проведения статистического исследования: наблюдений, группировок, анали-

за и обобщения первичных данных, выявления тенденций и закономерностей в развитии изучаемых объектов, явлений и процессов.

Качество разрабатываемой экономической политики определяется содержательным представлением о фактической ситуации в экономике страны, анализом возможностей и ограничений социально-экономического развития: ресурсных, структурных, научно-технических, социальных и др.

На разных уровнях управления и в различных отраслях и сферах деятельности существуют различные возможности и потребности в приемлемых методах анализа фактических данных и системах показателей, используемых для характеристики явлений общественной жизни.

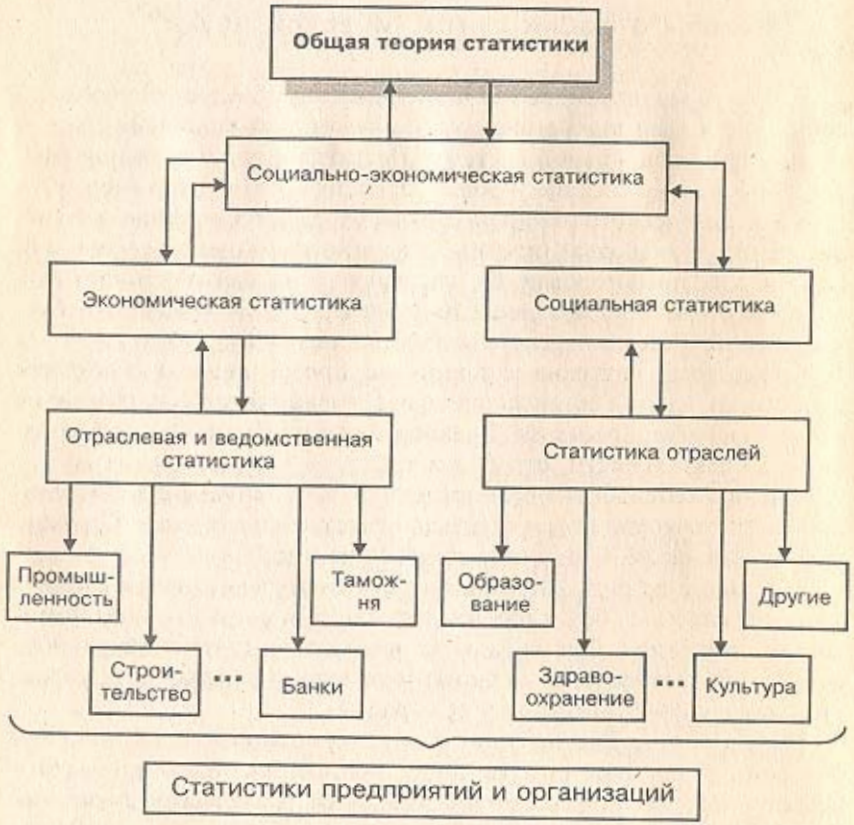


Рис. 1.1. Структура статистической науки

Все это находит свое отражение в содержании социально-экономической статистики и статистики отраслей. При этом экономическая статистика имеет своей задачей:

- разработку методологии статистического изучения наличия, движения и использования имеющихся ресурсов, обеспечивающих протекание изучаемых процессов;
- характеристику хода самого экономического процесса и его результатов;
- разработку системы обобщающих показателей и методики их расчета, в совокупности обеспечивающих количественную оценку результатов функционирования экономики страны и регионов, а также секторов, отраслей, отдельных предприятий и форм собственности.

Обобщающие показатели экономической статистики характеризуют состояние и тенденции развития национальной экономики, наличие, размещение и эффективность использования всех видов ресурсов (национального богатства, трудовых и финансовых ресурсов), состояние товарных рынков, позволяет проводить анализ инвестиций, инноваций и др.

Экономическая статистика раскрывает важнейшие пропорции и соотношения между производством, распределением, накоплением и потреблением.

Поскольку экономика, рассматриваемая как единый народнохозяйственный комплекс страны, представляет собой совокупность отдельных отраслей, то с целью получения сводных народнохозяйственных показателей в рамках экономической статистики разрабатываются общие принципы построения системы статистических показателей и методология их исчисления для всех отраслей народного хозяйства.

Для этого экономическая статистика использует данные статистики отдельных отраслей — промышленной, сельскохозяйственной, труда, финансовой, торговли, транспорта, внешнеэкономических связей и др.

Задачей социальной статистики является разработка и анализ системы показателей для характеристики качества и уровня жизни населения страны, отражающей демографическую ситуацию и различные аспекты социальных отношений. Ее отрасли — статистика управления, здравоохранения, образования, культуры, статистика населения, политическая статистика и др.

В рамках социально-экономической статистики особое место занимает система национального счетоводства. Система национальных счетов используется для описания и анализа развития рыночной экономики на макроуровне, характеризует производство, образование, распределение, перераспределение и использование

доходов, накопление национального богатства и связи страны с остальным миром.

Подобная информация необходима органам государственного управления для регулирования рыночной экономики, предпринимателям для оценки общей макроэкономической среды, в которой функционируют их предприятия, наконец, партнерам, организующим бизнес на межгосударственном уровне, поскольку представляет экономику страны на международном уровне.

Современная экономика России, ориентированная на развитие рыночных отношений, характеризуется не только изменением условий, а следовательно, и сути протекающих в обществе социально-экономических процессов, но и появлением целого ряда новых явлений, ранее не отслеживаемых отечественной статистикой.

Изучение этих социально-экономических процессов ставит перед статистикой задачу разработки методологии статистической оценки и анализа новых процессов и их результатов, совершенствования применяемых технологий в отношении традиционно изучаемых явлений. Проведение подобной работы является необходимым условием обеспечения субъектов всех уровней управления необходимой информацией для принятия обоснованных решений по регулированию и прогнозированию социально-экономического развития страны.

Глава 2
СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ

2.1. В чем заключаются требования, предъявляемые к статистической информации?

В процессе сбора и предоставления статистических данных органы государственной и ведомственной статистики выполняют роль посредников между респондентами (предприятиями, физическими лицами, общественными организациями и другими объектами наблюдения) и потребителями статистических данных. При этом направления информационных потоков можно представить в виде схемы (рис. 2.1).

Получение первичных данных является задачей статистического наблюдения. ***Статистическим наблюдением*** называется планомерный, научно обоснованный сбор данных или сведений о социально-экономических явлениях и процессах.

Создание информации, т.е. разработка агрегированных данных, которые относятся к большому числу отчетных единиц и представлены в виде таблиц или графиков, является задачей статистических расчетов производных показателей и анализа первичных данных.

В современном обществе, основными характеристиками которого являются динамичное развитие, появление новых видов деятельности, повышение деловой активности населения, развитие частного предпринимательства, активизация роли общественных и политических организаций, международная экономическая интеграция и т.д., становится трудно обозреть и систематизировать потребности государственных и негосударственных структур в информации. Вместе с тем эти обстоятельства повышают роль статистики в управлении по всей вертикали от макро- до микроуровня, поскольку основная задача статистики состоит в создании в обществе информационной инфраструктуры, которая должна обеспечить всем



Рис. 2.1. Посредническая роль статистики в создании информационных ресурсов

заинтересованным субъектам возможность для ориентации и анализа ситуации в экономике и социальной сфере.

Все эти явления обусловливают потребность в информации, которая:

1) шире по объему и тематически становится все более многообразной по мере развития экономического процесса;

2) обеспечивает базу как для большого числа сводных макроэкономических расчетов, так и для получения более дифференцированных структурных данных;

3) призвана не только предоставлять содержательную информацию, но и организовать рациональную систему сбора данных, не вызывающую больших затрат и использующую приемлемые для респондентов методы сбора данных.

Очевидно, что качественная информация опирается на достоверные, полные, сопоставимые и своевременные данные. Обеспечение достоверности данных является одной из важнейших задач, которую необходимо решать на стадии организации и проведения наблюдения. Достоверность данных зависит от следующих взаимодействующих факторов:

- независимости в области методологии и нейтральности официальной статистики по отношению к государственной власти;
- квалификации специалиста-статистика;
- качества используемого статистического инструментария;
- обеспечения конфиденциальности и защиты данных.

Квалификация специалиста-статистика определяется уровнем его профессиональной подготовки, навыками и способностями. Качество статистического инструментария обусловливается качеством программы наблюдения, бланков, анкет и инструкций по их заполнению. В условиях рыночной экономики интересы бизнеса требуют соблюдения «коммерческой тайны» в вопросах, касающихся производственной деятельности предприятий. В связи с этим получение достоверных данных возможно только на условиях обеспечения их защиты и конфиденциальности.

2.2. В чем заключаются программно-методологические и организационные вопросы статистического наблюдения?

Получение полной и качественной информации, на которую можно опираться в практической работе или научном исследовании, зависит от правильной организации и постановки статистического наблюдения. В связи с этим практическому проведению статистического наблюдения предшествует работа по составлению подробного плана, обеспечивающего методологическую и организационную стороны массового сбора первичных данных.

В программно-методологическую часть плана статистического наблюдения включаются вопросы определения цели, объекта, единицы и программы наблюдения.

При подготовке статистического наблюдения в первую очередь формулируется его цель. ***Цель наблюдения*** — это познавательная задача наблюдения, которая в общем виде сводится к информационному обеспечению процесса принятия решений. Так, например, при разработке экономической и социальной политики государства важным инструментом является информация о рынке труда. Для сбора данных о состоянии рынка труда организуются наблюдения по вопросам занятости, безработицы, заработной платы по профессиям, затрат на рабочую силу.

Цель определяет объект статистического наблюдения. ***Объектом статистического наблюдения*** называется совокупность единиц изучаемого явления, подлежащая статистическому наблюдению. Установить объект наблюдения — значит точно определить состав и границы изучаемой совокупности. При выборе объекта наблюдения необходимо четко определить его главные ***черты*** и ***свойства***, отличающие данный объект от других, смежных с ним. В противном случае можно получить неправильные выводы по сути изучаемого явления.

Важность правильного определения объекта наблюдения можно проиллюстрировать на следующем примере. Существует порог экономической безопасности по уровню безработицы в размере 6—8%, превышение которого ведет к социальной напряженности и социальным конфликтам в обществе. Это одна из причин, обусловливающих необходимость отслеживания и принятия меры к регулированию уровня занятости и безработицы. Что же в этом случае следует считать объектом наблюдения? В данном случае недостаточно указать, что объектом наблюдения является население, надо конкретизировать, что это экономически активное население, т.е. часть населения, которая предлагает свой труд для производства товаров и услуг. При проведении подобных обследований в России минимальный возраст для измерения экономической активности населения установлен в 15 лет, а максимальный — в 72 года. Для определения границ отнесения к категории экономически активного населения используют возрастные границы, т.е. ценз. ***Ценз*** — это ряд характерных признаков, при наличии которых явления и процессы относятся к изучаемой совокупности.

Объект наблюдения — это исследуемая статистическая совокупность, состоящая из отдельных единиц. Следовательно, такое

представление об исследуемой совокупности можно получить только при условии, что имеется информация, характеризующая ее отдельные единицы. В связи с этим наряду с установлением объекта наблюдения определяют единицу наблюдения и единицу совокупности.

Единица наблюдения является источником сведений, которые получают в результате наблюдения, а ***единица совокупности*** — носителем признаков, подлежащих наблюдению.

Так, например, при изучении структуры грузовых перевозок на автомобильном транспорте общего пользования по видам грузов объектом статистического исследования является совокупность крупных и средних автотранспортных предприятий, осуществляющих перевозки грузов на коммерческой основе (т.е. за плату); единицами наблюдения — конкретные предприятия; единицами совокупности — эксплуатируемые автомобили. Иногда единица наблюдения совпадает с единицей совокупности.

Объект наблюдения и его единицы характеризуются, как правило, множеством признаков различной степени значимости. В соответствии с целями и задачами наблюдения решается вопрос: какие именно признаки подлежат регистрации. Перечень таких признаков составляет ***программу наблюдения***.

Программа наблюдения оформляется в виде статистического бланка, состоящего, как правило, из титульной, адресной и содержательной частей. Статистический бланк может быть представлен в форме отчета, переписного листа, карточки учета, анкеты, группировочной таблицы.

В титульной части указывается наименование статистического органа, проводящего наблюдение, кем и когда утвержден формуляр, его название и номер.

В адресной части указывается адрес и наименование отчетной единицы, а для юридических лиц (если бланк имеет форму отчета) — отрасль, вид деятельности (код по ОКОНХ — Общероссийскому классификатору отраслей народного хозяйства), организационно-правовая форма (код по КОПФ — Классификатору организационно-правовых форм), орган управления государственным имуществом (код по ОКПО — Общероссийскому классификатору предприятий и организаций), единица измерения. Наличие кодов позволяет не только механизировать процесс обработки полученных в ходе наблюдения первичных данных, но и систематизировать их по отраслям, видам деятельности, организационно-правовым формам функционирования предприятий.

В содержательной части приводится перечень признаков, подлежащих регистрации, и соответствующая им информация.

Так, если обратиться к нашим примерам, то в случае наблюдения за перевозками грузов крупными и средними предприятиями автомобильного транспорта бланк наблюдения в его содержательной части состоит из двух разделов. Первый раздел содержит информацию о наличии на предприятии на начало обследуемого периода эксплуатационных автомобилей, сгруппированных по грузоподъемности, а также данные о распределении перевозок грузов и грузооборота за обследуемый период по грузоподъемности автомобилей. Во втором разделе содержится информация об объемах перевозок за обследуемый период по видам грузов и общем грузообороте по видам сообщения. Кроме того, отражаются сведения об общем пробеге грузовых автомобилей, в том числе с грузом.

В случае наблюдения за безработицей данные о безработице фиксируются по полу, возрасту, семейному положению; изучается распределение численности безработных по образованию, профессиональной принадлежности, регионам; учитывается продолжительность безработицы.

Для обеспечения единообразия в понимании вопросов и правильного заполнения бланков разрабатываются ***инструкции***. Вся документация по наблюдению (бланки, инструкции и т.д.) называется ***статистическим инструментарием***.

При проведении статистического наблюдения наряду с программно-методологической подготовкой большое значение имеет решение ***организационных вопросов***.

Круг этих вопросов достаточно широк и представлен в организационной части плана наблюдения, где указано, какие органы, в какие сроки и где проводят наблюдение, способы привлечения и подготовки квалифицированного персонала; источники обеспечения финансовыми, материальными и другими видами ресурсов, а также отражены специфические особенности организации конкретного наблюдения. Подготовка документации и ее рассылка республиканским, краевым, областным и другим комитетам и управлениям статистики также включаются в организационные вопросы. Так, например, при организации обследования населения по проблемам занятости в России Госкомстат РФ разрабатывает инструментарий обследования, организационные положения по его проведению. На региональном уровне территориальные органы государственной статистики составляют списки домашних хозяйств, включенных в выборку, формируют счетные участки, организуют оповещение населения о предстоящем обследовании, проводят подбор и обучение интервьюеров (счетчиков); обеспечивают интервьюеров пособиями по организации и проведению опроса бланками анкет и инструкциями по их заполнению, а также осуществляют выборочную проверку работы интервьюеров путем установ-

ления повторного контакта с опрошенными лицами и первичного контроля анкет обследования. Работа по организации проведения обследования населения осуществляется во взаимодействии с органами правопорядка и службы занятости. Таким образом, решение программно-методических и организационных вопросов статистического наблюдения является необходимым условием его успешного проведения.

2.3. Каковы организационные формы статистического наблюдения?

В российской статистике традиционно сложились две основные организационные формы статистического наблюдения:

- отчетность;
- специально организованные наблюдения.

Отчетность – это статистическое наблюдение, при котором органы государственной статистики получают сведения о деятельности предприятий, учреждений, организаций путем представления ими данных в специальных документах, называемых формами. Формы отчетности представляются в строго установленные сроки.

Каждая форма отчетности имеет определенный номер, название с указанием периодичности или даты представления, реквизиты предприятия (организаций), его административно-территориальное нахождение, коды по классификаторам и сведения по специальной программе. Периодичность представления отчетности может быть различной: месячной, квартальной, полугодовой, годовой.

Статистическая отчетность является обязательной и имеет юридическую силу, руководитель организации, подписывающий формы отчетности, несет ответственность за достоверность представляемых данных. Сведения форм статистической отчетности базируются на данных первичного учета.

В условиях административной экономики отчетность была всеобъемлющей и включала большое число форм с огромным числом показателей. В настоящее время в связи с высокой стоимостью информационных ресурсов, а также появлением понятия «коммерческая тайна» оказалось целесообразным не только сокращение количества форм наблюдения, но и их унификация.

Госкомстатом РФ утверждены следующие унифицированные формы статистического наблюдения (табл. 2.1).

Использование унифицированных форм статистического наблюдения позволяет получить комплексную информацию по всем ви-

Таблица 2.1

Унифицированные формы статистического наблюдения

№ п/п	Номер формы	Название	Периодичность представления
1	Форма №1 — предприятие	Основные сведения о деятельности предприятия	Годовая
2	Форма № П-1	Сведения о производстве и отгрузке товаров и услуг	Месячная
3	Форма № П-2	Сведения об инвестициях	Квартальная
4	Форма № П-3	Сведения о финансовом состоянии организации	Месячная
5	Форма № П-4	Сведения о численности, заработной плате и движении работников	Месячная
6	Форма № П-5	Основные сведения о деятельности малого предприятия	Квартальная

дам деятельности предприятия и о его финансовом положении, к какой бы отрасли оно ни относилось.

Важным условием организации наблюдения на основе унифицированных форм является разработка и утверждение форм первичной учетной документации, адаптированных к требованиям международных стандартов бухгалтерского учета, а также их электронных версий. Госкомстатом РФ разработан альбом унифицированных форм первичного учета, включающий формы по учету труда и его оплате, основных средств и нематериальных активов, материалов, малоценных и быстроизнашивающихся предметов, работ в капитальном строительстве и др.

Внедрение унифицированных форм первичной учетной документации способствует повышению качества, достоверности и сопоставимости результатов наблюдения.

В дополнение к унифицированным формам используются формы наблюдения, отражающие специфическую информацию отраслевого характера. Это так называемая ***специализированная отчетность***.

Имеются специализированные формы для наблюдения на транспорте, в связи, торговле, строительстве, для предприятий естественных монополистов и др. В целях усиления контроля за увеличением сбора налогов и сокращения расходов федерального бюджета с 1 января 1997 г. введена месячная форма государственного статистического наблюдения о составе кредиторской задолженнос-

ти организаций-бюджетополучателей. Для наблюдения за деятельностью финансово-промышленных групп (ФПГ) введена форма наблюдения №1 — ФПГ, данные которой позволяют охарактеризовать ФПГ, оценить структуру по видам деятельности, ресурсный потенциал, масштабы производства продукции и инвестиционной деятельности.

В современных условиях особая роль принадлежит второй форме — ***специально организованным наблюдениям***, поскольку в ряде случаев только такого рода наблюдения позволяют получить данные для управления социально-экономическими процессами. Специально организованные наблюдения проводятся в виде ***переписей*** или различного рода ***обследований***. Одной из важнейших работ такого рода, выполняемых государственной статистикой, является перепись населения.

Переписи населения широко используются в мировой статистической практике. В России перепись населения за последнее столетие проводилась девять раз (1897, 1926, 1936, 1937, 1939, 1959, 1970, 1979, 1989 гг.). Периодичность ее проведения нарушалась особыми условиями (Гражданская война, период сталинских репрессий, Великая Отечественная война). Перепись населения дает сведения о численности населения, его национальном и социальном составе, миграции, занятости, образовании, рождаемости и смертности, характеристике жилищных условий, заболеваемости и др., что позволяет судить о социально-демографических процессах в обществе.

Очередная перепись населения России должна состояться в октябре 2002 г. В этой связи в Правительство РФ внесен проект закона «О всероссийской переписи населения», в соответствии с которым должна быть проведена большая организационная и методическая работа, обеспечивающая успешность ее проведения.

Программа предстоящей переписи учитывает специфику современных условий России и соответствует международным рекомендациям, при этом обеспечивается возможность получения данных, как сравнимых с прошлыми переписями, так и применимых в международных сопоставлениях.

Основные особенности предстоящей переписи будут состоять в том, что в качестве единицы наблюдения должно быть использовано домохозяйство (а не семья, как в предыдущих). Это позволит выполнить более глубокий анализ структуры как домохозяйства, так и семьи, обеспечит возможность сравнительного анализа с данными переписей других стран.

Существенным изменением является включение в программу ряда новых вопросов, в частности, связанных с изучением занятости, что позволит выявить структуру рабочей силы, число безработных,

их социальный состав и демографическую характеристику, а это, в свою очередь, даст возможность определить экономически активное и неактивное население.

Важным будет блок вопросов о миграции населения, обработка которых позволит получить полную характеристику ее величины, направлений передвижения, состава мигрантов, определить численность беженцев и вынужденных переселенцев. Новыми являются вопросы, связанные с изучением числа браков и рождаемости, ответы на которые позволят сопоставить данные переписи и органов загса.

С 10 по 17 октября 2000 г. будет проведена пробная перепись населения в Москве, Московской области и Владивостоке с общим охватом 100 тыс. человек, в ходе которой будет отработан комплекс вопросов методологического и организационного характера.

С целью изучения отдельных вопросов проводятся переписи в отраслях экономики — промышленности, строительстве, сельском хозяйстве, торговле и т.д. Так, с целью получения достоверных данных о производстве продукции малыми предприятиями Госкомстат РФ планирует провести сплошное обследование малых предприятий по итогам работы за 2000 г.

Регулярно, начиная с 1992 г., проводятся обследования с целью изучения занятости населения, динамики цен в отраслях экономики, потребительских цен и др.

Нетрадиционным для российской статистики является способ получения информации на основе ***бизнес-обследований***, достаточно широко применяемый в международной практике в течение последних десятилетий.

Проведение бизнес-обследований хозяйственной деятельности строится на сборе и обобщении информации, основанной на оценке текущего экономического состояния и ожидаемого изменения основных показателей хозяйственной деятельности. Респонденты также отвечают на вопросы, позволяющие выявить факторы, влияющие на деятельность их структур, и ряд других «качественных» вопросов.

В России бизнес-обследования проводятся с 1992 г. В настоящее время бизнес-обследования проводятся статистическими органами России совместно с Центром экономической конъюнктуры при Правительстве Российской Федерации. Ежемесячно обследуются предпринимательские намерения (деловая активность) в промышленности, строительстве, сельском хозяйстве, оптовой торговле, инновационной, банковской и страховой сферах. Важной целью их проведения является совершенствование экономического анализа и прогнозирование. Динамические ряды, полученные в результате обследований, позволяют определять поворотные точки в экономике и отслеживать циклические процессы.

Необходимость постоянного отслеживания динамически меняющихся процессов в социальной и экономической сферах привела к созданию ***специальных наблюдений — мониторингов***. Так, например, ведутся мониторинги социально-трудовой сферы, основными направлениями которых являются: новые социально-трудовые процессы; доходы и уровень жизни населения; состояние условий труда; использование рабочего времени; занятость. Ценность мониторинга заключается не столько в характеристике этих процессов, сколько в возможности на основе оперативных данных выявлять негативные тенденции и факторы, их вызывающие, а также оценивать эффективность и полноту реализуемых законов и нормативных актов по социально-трудовым вопросам.

Специфическими наблюдениями являются так называемые ***пилотные обследования***. Так, в 2000 г. будет проведено пилотное обследование индивидуальных предпринимателей, осуществляющих розничную торговлю, что позволит уточнить показатели торгового оборота по субъектам малого предпринимательства.

Таким образом, используемые российской статистикой различные формы статистического наблюдения позволяют получить обширную информацию для управления социально-экономическими процессами в стране.

2.4. Какие виды статистического наблюдения существуют?

Статистическое наблюдение делится на виды по времени проведения, охвату единиц совокупности, источникам получения первичных данных (рис. 2.2).

По времени проведения наблюдение бывает ***непрерывным***, или текущим, ***периодическим*** и ***единовременным***. При непрерывном наблюдении изменения в социально-экономических явлениях регистрируются по мере их свершения. Примером такого наблюдения является запись актов гражданского состояния: рождение, смерть, состояние в браке.

Периодические наблюдения проводятся через определенные периоды времени по схожим программе и инструментарию. Таким видом наблюдений может служить регистрация потребительских цен на товары и услуги, осуществляемая каждый месяц.

Единовременное наблюдение проводится один раз для получения данных по определенному кругу вопросов или повторяется через неопределенные промежутки времени по мере возникновения потребности в данных. Примером единовременного наблюдения являются переписи населения, производственных мощностей, жилого фонда и т.д.

По охвату единиц совокупности наблюдение делится на ***сплошное*** и ***несплошное***. При проведении сплошного наблюдения регистрируются все без исключения единицы совокупности. Примером сплошного наблюдения является перепись организаций оптовой торговли, которая охватывает всю страну, включая городские и сельские районы. В перепись включаются все организации независимо от размера и форм собственности. В процессе переписи необходимо получить данные о количестве действующих предприятий и изучить процессы, происходящие в оптовой торговле. Целью переписи является получение данных о числе действующих организаций, количественной информации о развитии оптовой торговли в стране, влиянии оптовой торговли в формировании макроэкономических показателей. Использование сплошного учета является основой для разработки регистра организаций оптовой торговли. Единицей наблюдения переписи является организация оптовой торговли, состоящая на самостоятельном балансе, независимо от организационно-правовых форм и форм собственности, осуществляющая оптовую продажу или оказывающая оптово-посредническое услуги по купле-продаже продукции.

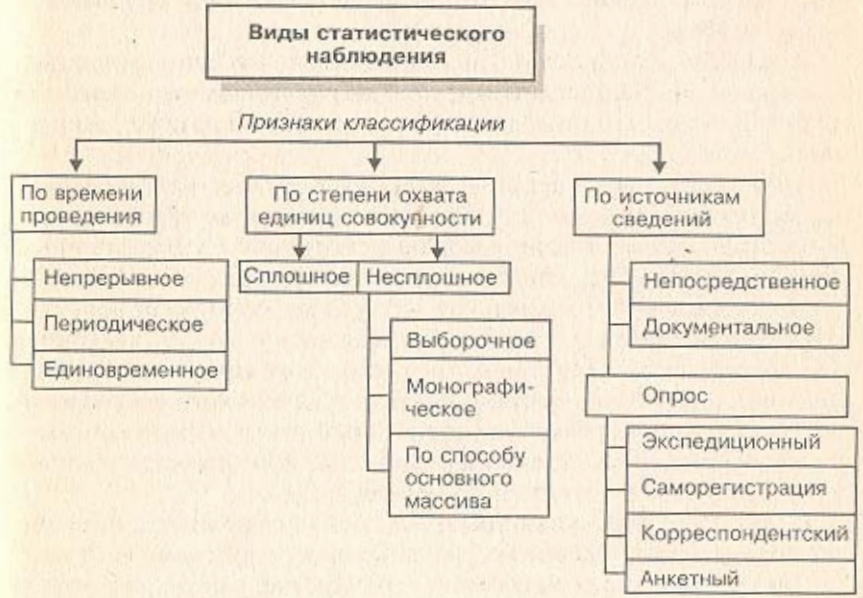


Рис. 2.2. Виды статистического наблюдения

В последнее время все более широкое распространение получают ***несплошные наблюдения***, которые могут быть организованы как выборочные, способ основного массива и монографические.

При ***выборочном наблюдении*** обследованию подвергается отобранная в определенном порядке часть единиц совокупности, а полученные результаты распространяются на всю совокупность.

Отбор производится таким образом, чтобы выборочная совокупность была репрезентативна, т.е. как можно точнее воспроизводила всю совокупность со всеми характерными для нее особенностями.

При использовании ***способа основного массива*** обследованию подлежат наиболее существенные крупные единицы, которые имеют максимальный удельный вес в совокупности. Такое наблюдение используется для получения данных о ценах на товары, реализуемых на рынках. Например, статистическое наблюдение за изменениями цен производителей промышленной продукции осуществляется органами статистики по выборочной сети базовых предприятий, отбор которых проводится в два этапа:

отбирается минимум крупных предприятий, имеющих значительный вес в промышленности Российской Федерации, т.е. таких предприятий, на которых стоимость отобранных товаров составляет не менее 50% стоимости выпуска;

региональные статистические органы дополнительно вносят в этот перечень предприятия, отражающие специфику промышленности региона.

В отличие от выборки и способа основного массива, являющихся массовыми обследованиями, при ***монографическом наблюдении*** тщательному изучению подвергаются отдельные единицы совокупности.

Цель такого наблюдения — детальное описание каждой единицы, которое невозможно при проведении массовых обследований. Монографическое описание можно использовать для характеристики отдельного предприятия, бюджета семьи и др.

Массовое и монографическое наблюдения дополняют друг друга. С одной стороны, для того чтобы включить ту или иную единицу в монографическое описание, требуются данные массовых обследований, с другой — результаты монографического наблюдения могут быть использованы для составления программы нового массового обследования, учета специфических особенностей и характерных признаков объекта наблюдения.

В зависимости от источников получения первичных данных различают непосредственное наблюдение, документальное и опрос.

Непосредственное наблюдение — это наблюдение, при котором источником записей, производимых регистратором в статистический бланк, является непосредственный счет единиц наблюдения в

натуре либо измерение их признаков в результате осмотра, обмера, взвешивания. Примером могут служить снятие остатков товарно-материальных ценностей при проведении их инвентаризации; хронометражные наблюдения затрат рабочего времени на выполнение трудовых операций в процессе разработки нормированных заданий, фиксация потерь рабочего времени при контроле за использованием внутрисменного рабочего времени в ходе моментных наблюдений.

Документальное наблюдение основано на использовании в качестве источника первичных данных сведений из различных документов. Так, при проверке хозяйственных операций по существу, т.е. с целью установления ошибок или злоупотреблений при подготовке представленной для подтверждения бухгалтерской отчетности, аудитор проводит выборочную проверку документов первичного учета предприятий, организаций, учреждений.

При ***опросе*** сведения о каждой единице наблюдения фиксируются со слов опрашиваемых в бланках анкет. Такой вид наблюдения характерен для переписей населения, проведения социологических опросов, маркетинговых исследований по вопросам выяснения предпочтений потребителей и оценки ими свойств товаров. В статистической практике применяются следующие виды опросов: экспедиционный, саморегистрация, корреспондентский.

При ***экспедиционном способе*** с каждой единицей наблюдения работают специально подготовленные люди, называемые счетчиками или регистраторами, которые получают сведения со слов опрашиваемых и сами фиксируют их в специальных формулярах. Такой способ используется при сборе данных у населения, в частности, при переписях.

При ***саморегистрации*** специальные работники раздают опрашиваемым бланки и дают инструкции по их заполнению; заполняют же бланки сами респонденты. Этот способ лежит в основе, например, бизнес-обследований. В хода бизнес-обследований собирается и обобщается «качественная» информация, получаемая от топ-менеджеров, по оценке текущего экономического состояния структур и намерений на ближайшую перспективу. Респондентам предлагается также оценить фактическое и ожидаемое изменение основных показателей производственно-финансовой деятельности в рамках анализа тенденций «увеличение — уменьшение», «улучшение — ухудшение», «осталось без изменения». Кроме того, в программу обследования входят вопросы, связанные с выявлением факторов, ограничивающих деятельность экономических единиц. Результаты бизнес-обследований являются информационной базой для анализа и прогнозирования состояния экономики в целом и отдельных ее секторов, для оценки изменений экономической конъюнктуры.

Корреспондентский способ основан на том, что сведения в органы статистики сообщает штат добровольных корреспондентов. Этот способ применяется, в частности, при обследовании доходов домашних хозяйств.

Анкетный способ состоит в том, что разработанная анкета рассылается определенному кругу лиц и после заполнения возвращается статистическому органу или другим организациям, проводившим исследование.

При проведении анкетирования следует учитывать, что, как правило, возвращается статистическому органу не более 80% разосланных анкет, поскольку не все респонденты заинтересованы в опросе, кроме того, часто возникают трудности в заполнении анкет.

Поэтому при проведении анкетирования следует тщательно разработать его инструментарий, т.е. подготовить инструкции по заполнению анкет. Саму анкету следует разрабатывать в соответствии со следующими требованиями:

- должна быть четко определена и сформулирована цель исследования;
- вопросы не могут быть двусмысленными и должны соответствовать уровню образованности потенциальных респондентов;
- вопросы следует формулировать корректно, так, чтобы они не вели к возрастанию коэффициента *Non-Response-Rate* («неполучение ответов»);
- формулировать вопросы нужно по возможности нейтрально, т.е. избегать формулировок «за», «против»;
- иногда целесообразно приводить так называемый «подсказ».

На практике форма наблюдения, вид наблюдения по времени проведения, охвату единиц совокупности, а также способ регистрации первичных данных устанавливаются исходя из задачи наблюдения, специфики наблюдаемого объекта, срочности представления результатов и обеспечения требуемой степени их точности, а также материальных возможностей заказчика (потребителя) информации.

2.5. В чем заключаются основные направления совершенствования статистического наблюдения в современных условиях?

Совершенствование организации статистического наблюдения на современном этапе осуществляется на основе положений федеральной программы реформирования статистики. Прежде всего это переход от отраслевого принципа получения статистической

информации к статистике предприятий, что предполагает разработку и ведение Единого регистра предприятий и организаций (ЕГРПО). В регистре по каждому субъекту содержатся сведения о его юридическом и фактическом положении, отраслевой принадлежности, основных экономических характеристиках (объеме продукции, стоимости основных средств, численности работающих и др.). На базе ЕГРПО ведутся различные субрегистры (малых предприятий, строительных организаций, организаций розничной и оптовой торговли и др.). На основе субрегистров формируется генеральная совокупность объектов статистического наблюдения, что, в свою очередь, позволяет формировать выборочные совокупности. Данные субрегистров постоянно уточняются на основе бухгалтерской отчетности, данных Министерства по налогам и сборам, Министерства внутренних дел и др.

Создается регистр населения, который ведется по регионам страны и содержит сведения о каждом жителе (идентификационный номер, фамилия, имя, отчество, дата и место рождения, пол, национальность, гражданство, адрес постоянного места жительства и др.). Таким образом, регистры служат источником статистических данных.

Другим направлением совершенствования статистического наблюдения является сокращение отчетности и переход на унифицированные формы текущего статистического наблюдения, что подробно рассмотрено ранее.

Важным направлением является использование сплошного наблюдения только по крупным и средним предприятиям и выборочного по малым, что приводит к значительной экономии государственных ресурсов. Такая система принята во всех странах с рыночной экономикой.

Расширение масштабов выборочного наблюдения предполагает дальнейшее совершенствование его методологии и практической организации (см. главу 6).

Особое внимание уделяется не только совершенствованию традиционных методов статистического наблюдения (переписей, выборочных обследований), но и использованию новых методов (бизнес-обследований, пилотных, цензовых обследований и др.), проведение мониторинга по важнейшим направлениям экономических и социальных процессов, а также использование сочетания этих методов.

Так, например, получение достоверных сведений о деятельности малых предприятий достигается на основе сочетания квартальных выборочных наблюдений со сплошными обследованиями. Во многих регионах и на федеральном уровне действуют системы мониторинга оценки динамики социальных, экономических и экологических процессов.

Проводится пилотное обследование образования, использования, обезвреживания и размещения отходов производства и потребления на территории Мурманской области (совместно с ЦСБ Финляндии).

Принципиально новым для российской государственной статистики является использование научных дорасчетов социально-экономических показателей. В настоящее время уже разработаны методики дорасчетов показателей СНС, объемов продукции в строительстве, на транспорте, в сельском хозяйстве, товарооборота в торговле и др. Существуют методики определения скрытой неформальной деятельности, которая составляет 20 — 25%, а по данным некоторых экспертов — до 70% ВВП.

Решение стоящих перед статистикой задач невозможно без принципиального изменения самого процесса получения данных на базе современных средств вычислительной техники и информационных технологий, позволяющих создать интегрированную базу данных.

Большой объем информации содержит отчетность учреждений государственного управления, государственных внебюджетных фондов, различных ведомств. Значительная роль в информационном обеспечении принадлежит банковской, таможенной, налоговой статистике и другим, так называемым ведомственным, статистикам. Только сочетание всех источников информации может дать объективную картину социально-экономических процессов, что необходимо для их государственного регулирования.

2.6. В чем заключаются основные ошибки статистического наблюдения и каковы способы контроля статистических данных?

Ошибкой наблюдения считается величина отклонения между расчетным и фактическим значениями по изучаемым признакам. В зависимости от причин возникновения выделяются ошибки репрезентативности и ошибки регистрации.

Ошибки репрезентативности — это расхождение между значением признака по результатам наблюдения и действительным его значением. ***Ошибки регистрации*** возникают при сплошном и несплошном видах наблюдения. Ошибки регистрации могут быть преднамеренными и непреднамеренными.

Преднамеренные ошибки могут существенно отразиться на качестве информации. Решение этой проблемы возможно в определенной мере в результате реализации механизмов, направленных на обеспечение информационной безопасности статистической системы. В России получило распространение занижение размеров прибыли, указанных в отчетности. Поэтому обеспечение конфиденциальности представляемых данных позволит устранить преднамеренные ошибки и тем самым повысить качество информации. Среди ***непреднамеренных ошибок*** различают случайные и систематические ошибки.

Случайные ошибки возникают в результате влияния различных случайных факторов. Они разнонаправленны, т.е. могут увеличивать и уменьшать значение показателя. При достаточно большом объеме совокупности в соответствии с действием закона больших чисел эти ошибки взаимопогашаются.

Систематические ошибки могут привести или к увеличению, или к уменьшению значения по отдельным единицам наблюдения. Следовательно, полученные по таким данным на стадии обобщения показатели будут содержать накопленную ошибку. Источником возникновения подобных ошибок может являться:

- отсутствие данных по некоторым единицам совокупности, что может возникнуть вследствие невнимательного заполнения бланков или отказа респондентов представить сведения;
- неправильное заполнение отчетности, бланков, а при обследовании населения — неправильные ответы;
- нарушение методологии сбора данных;
- неточности при проверке записей, кодировании и предоставлении результатов в таблице.

Из перечисленных источников ошибок особое внимание необходимо обратить на те из них, которые связаны с неполучением ответов или невозможностью их получения. Примерные ситуации невозможности получения ответов сводятся к следующим:

- необнаруженные, т.е. те единицы, которые оказалось невозможным найти или посетить, например, при территориальном наблюдении;
- неспособность дать ответы на вопросы. Мерой предосторожности против этих ошибок является тщательная формулировка вопросов в опросном листе и его предварительное испытание;
- «крепкие орешки», т.е. лица, которые категорически отказываются отвечать. Ошибки такого рода устранить практически невозможно.

Для выборочного наблюдения характерны ошибки репрезентативности, связанные с различиями в составе генеральной и обследуемой совокупностей. Методы определения ошибок репрезентативности рассмотрены в гл. 6.

После получения формуляров необходимо проверить их количество и правильность записи в них данных. Иногда об отсутствии единиц указывает сравнение подсчетов их числа с другими аналогичными подсчетами при другом подобном обследовании. Также для проверки можно использовать данные регистра.

Следующим этапом является арифметический контроль, который основан на количественной взаимосвязи показателей. Кроме арифметического проводится логический контроль, осуществляемый на основе логических взаимосвязей показателей. Для исправления ошибок, обнаруженных в результате арифметического и логического контроля, необходимо получить повторное подтверждение у источника представления данных.

Глава 3
ГРУППИРОВКА СТАТИСТИЧЕСКИХ ДАННЫХ И ЕЕ РОЛЬ В АНАЛИЗЕ ИНФОРМАЦИИ

3.1. В чем состоит значение метода группировок в анализе статистических данных?

Одним из основных и наиболее распространенных методов обработки и анализа первичной статистической информации является группировка. Понятие статистической группировки в широком смысле слова охватывает целый комплекс статистических операций. Прежде всего к ним относится объединение зарегистрированных при наблюдении единичных случаев в группы, сходные в том или ином отношении, поскольку целостную характеристику совокупности необходимо сочетать с характеристикой основных ее частей, классов и т.д. Затем проводится подсчет итогов по выделенным группам и всей совокупности в целом и, наконец, результаты группировки оформляются в виде ***статистических таблиц***. При составлении плана статистического наблюдения определяется очередность обработки материалов, разрабатываются макеты сводных таблиц, указывается также, кто и в какие сроки осуществляет сводку, каким способом, куда поступают сводные данные и кто проводит их дальнейшую обработку.

Значение группировок состоит в том, что этот метод, во-первых, обеспечивает систематизацию и обобщение результатов наблюдения, а во-вторых, метод группировок является основой применения других методов статистического анализа основных сторон и характерных особенностей изучаемых явлений.

В процессе группировки происходит распределение единиц статистической совокупности по группам, однородным в каком-либо существенном отношении, и дается характеристика таких групп с помощью системы показателей в целях выделения типов явлений, изучения их структуры и взаимосвязей. Метод группировок выполняет некоторые функции, аналогичные функциям эксперимента в естественных науках: посредством группировки по отдельным признакам и комбинации самих признаков статистика имеет возможность выявить закономерности и взаимосвязи явлений в условиях, в известной мере ею определяемых. При использовании метода груп-

пировок появляется возможность проследить взаимоотношения различных факторов и определить силу их влияния на результативные показатели.

Огромное значение и роль группировок в статистическом исследовании вытекает из характера объекта статистики, его специфики. Явления общественной жизни, изучаемые статистикой, отличаются многообразием форм и стадий развития, они состоят из существенно различающихся частей, обладающих многими специфическими свойствами.

Изучая количественную сторону массовых общественных явлений в неразрывной связи с их количественными особенностями, статистика стремится показать совокупность явлений дифференциации в многообразии их типов, рассмотреть взаимосвязи и соотношения между последними. С помощью метода группировок решаются сложные задачи статистического анализа.

3.2. Какие виды группировок следует выделить?

Необходимость группировки обусловливается прежде всего наличием качественных различий между изучаемыми явлениями, и первую задачу группировок можно сформулировать как задачу выделения в составе массового явления тех его частей, которые однородны по качеству и условиям развития, в которых действуют одни и те же закономерности влияния факторов. В результате такой группировки выделяются социально-экономические типы (а отсюда и название группировки — ***типологическая***) как выражение конкретного общественного процесса, его форм и разветвлений, как выражение существенных черт, общих для множества единичных явлений.

С задачами типологической группировки тесно связаны и две другие задачи группировок: характеристика структуры и структурных сдвигов в исследуемой совокупности и выявление взаимосвязи отдельных признаков изучаемого явления.

Примерами типологических группировок могут служить группировки хозяйственных объектов по формам собственности, разделение экономически активного населения на занятых и безработных, работников на занятых преимущественно физическим и умственным трудом и т.д.

Методология типологических группировок определяется тем, насколько ясно проявляются качественные различия изучаемых явлений. Например, при группировках отраслей промышленности по экономическому назначению продукции выделяются отрасли,

производящие средства производства, и отрасли, производящие предметы потребления, в макроструктуре розничного товарооборота выделяются производственные и непроизводственные товары. В большинстве случаев качественные различия между явлениями не проступают столь отчетливо. Например, выделение в отраслях промышленности крупных, средних и мелких предприятий является достаточно сложной в методологическом отношении проблемой. В подобных случаях после предварительного определения возможных типов на основе четкой формулировки познавательной задачи необходимо выявлять так называемые группировочные признаки.

Выбор группировочных признаков всегда должен быть основан на анализе качественной природы исследуемого явления. Всесторонний теоретико-экономический анализ сущности и закономерности развития явления должен быть направлен на то, чтобы в соответствии с целью и задачами исследования положить в основу группировки существенные признаки. В зависимости от конкретных условий группировочные признаки должны видоизменяться, т.е. должны быть специализированы при выделении одного и того же типа в различных условиях. Так, в настоящее время в промышленности при группировке предприятий по размерам могут использоваться показатели производственной мощности предприятий, стоимости промышленно-производственных основных фондов, численности промышленно-производственного персонала, объема потребляемой энергии. Если для трудоемких отраслей в качестве группировочного признака целесообразно использовать численность работников, то для фондоемких отраслей — стоимость промышленно-производственных фондов, для энергоемких отраслей — количество потребляемой энергии.

Вместе с тем следует учитывать, что полную характеристику каждому типу явлений можно дать при использовании ***системы признаков*** (***системы показателей***). Только совокупность признаков позволяет отобразить процессы развития, всесторонне выявить реальные связи, взаимоотношения отдельных сторон процесса.

Например, коммерческие банки обязаны создавать резерв для возмещения возможных потери по ссудам в соответствии со степенью риска погашения ссуды. В зависимости от степени риска все ссуды делятся на пять групп: стандартные, нестандартные, сомнительные, опасные и безнадежные. Для отнесения ссуды к той или иной группе риска используют группировку выданных банком ссуд по двум признакам: во-первых, в зависимости от наличия реального обеспечения (залог, гарантия банка, гарантия Правительства Российской Федерации), во-вторых, в зависимости от количества дней просрочки выплаты ссуды.

По степени обеспеченности ссуды подразделяются на три группы: 1) обеспеченные ссуды; 2) недостаточно обеспеченные ссуды; 3) необеспеченные ссуды.

К обеспеченным относятся ссуды, имеющие ликвидный залог, стоимость которого равна 100% размера ссуды, либо банковскую гарантию или гарантию правительства. Недостаточно обеспеченная ссуда имеет частичное обеспечение по стоимости не менее 60% от размера ссуды. Необеспеченная ссуда не имеет обеспечения либо обеспечение составляет менее 60% размера ссуды.

По количеству дней просроченной задолженности выделяют пять групп: 1) возврат ссуд в срок; 2) до 30 дней; 3) от 30 до 60 дней; 4) от 60 до 90 дней; 5) более 90 дней.

Результаты распределения всех кредитов по группам риска по состоянию на конец июля 1995 г. по одному из московских банков приведены в табл. 3.1 (общая сумма кредитного портфеля — 140 млн руб.).

Группы риска по ссудам формируются по сочетанию двух признаков.

Так, к первой группе риска — «Стандартные ссуды» — относятся ссуды, по которым своевременно и в полном объеме погашается основной долг, а также просроченные до 30 дней обеспеченные ссуды. В нашем примере «Стандартные ссуды» составляют 98 млн руб., или 70% общей суммы выданных кредитов.

Ко второй группе риска — «Нестандартные ссуды» — относятся просроченные до 30 дней недостаточно обеспеченные ссуды, а также просроченные от 30 до 60 дней обеспеченные ссуды. Таких ссуд насчитывается 14 млн руб., или 10% общей суммы выданных кредитов.

К третьей группе — «Сомнительные ссуды» — относятся просроченные до 30 дней необеспеченные ссуды, а также просроченные от 60 до 90 дней недостаточно обеспеченные ссуды. Эти ссуды составляют 14 млн руб., или 10% общей суммы выданных кредитов.

В четвертую группу — «Опасные ссуды» — включают все просроченные от 30 до 60 дней необеспеченные ссуды, а также просроченные от 60 до 90 дней недостаточно обеспеченные ссуды. Такие ссуды в примере составляют 9 млн руб., или 6,4% общей суммы выданных кредитов.

К пятой группе риска — «Безнадежные ссуды» — относятся просроченные от 60 до 90 дней необеспеченные ссуды и все ссуды, просроченные свыше 90 дней. Сумма ссуд в этой группе риска, по приведенным данным, составила 5 млн руб., или 3,6% общей суммы выданных кредитов.

Поскольку для каждой группы риска устанавливается определенный процент резервирования средств, приведенные данные в

Таблица 3.1

Распределение кредитного портфеля банка по группам риска на конец июля 1995 г. (в млн руб.)

Количество дней просроченной задолженности	Обеспеченность ссуды	Группы риска по ссудам					Итог по группе
		Стандартные	Нестандартные	Сомнительные	Опасные	Безнадежные	
1	2	3	4	5	6	7	8
Возврат ссуды в срок	Обеспеченная	62					80
	Недостаточно обеспеченная	11					
	Необеспеченная	7					
До 30 дней	Обеспеченная	18					42
	Недостаточно обеспеченная		4				
	Необеспеченная			2			
30—60	Обеспеченная		10				15
	Недостаточно обеспеченная			3			
	Необеспеченная				2		
60—90	Обеспеченная			9			17
	Недостаточно обеспеченная				7		
	Необеспеченная					1	
Более 90	Обеспеченная					1	4
	Недостаточно обеспеченная					2	
	Необеспеченная						
ИТОГО в группе		98	14	14	9	5	140

итоговой строке табл. 3.1 служат основой для определения необходимого резерва на случай непогашения ссуд.

Если в основу группировки положено несколько признаков, такую группировку называют ***сложной***. Сложная группировка может выполняться как комбинационная и как многомерная. При использовании ***комбинационной группировки*** (как в приведенном в табл. 3.1 примере) группы, выделенные по одному из признаков, затем подразделяются на подгруппы по другому признаку, в свою очередь подгруппы могут быть разделены по следующему признаку и т.д. Общее число выделенных групп будет равно произведению числа группировочных признаков на число выделенных групп по каждому из них. В случае, если группировка осуществляется не последовательно по отдельным признакам, а одновременно по комплексу признаков, ее называют ***многомерной*** (подробнее см. параграф 3.3).

3.3. Какие группировки называют классификациями?

В зависимости от группировочных признаков различают ***группировки*** по количественным и качественным признакам.

Качественный признак отражает определенные свойства качества данного явления и записывается в виде текста. Если качественный признак имеет мало разновидностей, то количество групп определяется числом этих разновидностей. Таковы, например, группировки населения по семейному положению, образованию, деление населения на городское и сельское и т.д.

Но нередки случаи, когда качественный признак имеет большое число разновидностей и перечислить их все не представляется целесообразным, например, профессии рабочих, номенклатура выпускаемой продукции, виды основных фондов и т.п. В таких случаях разрабатывают ***классификацию разновидностей,*** т.е. сходные по своим особенностям разновидности объединяются в группы (классы). Под классификацией обычно понимается более устойчивое разграничение единиц наблюдения, чем при группировке. Классификации используются в течение длительного времени, хотя со временем они могут подвергаться более или менее существенным изменениям. Утверждаются классификации, как правило, в качестве национального или международного стандарта. Широко используются классификации отраслей экономики, видов экономической деятельности, классификация продукции, работ и услуг и многие другие. Идентификация объектов осуществляется на основе

их кодирования, для чего разрабатывается соответствующий классификатор.

Любая классификация может состоять из нескольких иерархических уровней, т.е. выделенная на первом этапе классификационная группа может в дальнейшем детализироваться и т.д. Необходимость и степень детализации классификационных групп определяется целью и задачами статистического исследования.

С 1991 г. учет экспорта и импорта основывается на базе Гармонизированной системы описания и кодирования товаров (ГС) и имеет шесть ступеней классификации: разделы, группы, подгруппы, товарные позиции, подпозиции и субпозиции. В составе 21 раздела выделяется 96 групп по принципу выделения последовательных стадий обработки товаров: сырье — полуфабрикаты — готовые изделия. При построении 33 товарных подгрупп, 1241 позиции и 5019 субпозиций на каждом уровне применяется своя последовательность группировочных признаков, среди которых основными являются: назначение товар, степень его обработки, вид материала, из которого он изготовлен, удельный вес товара в оптовой торговле. Товарные разделы Гармонизированной системы описания и кодирования товаров приведены в Приложении 9.

Товарная классификация является ресурсом упорядочения информации о внеэкономических связях, служит основой их системного анализа и обеспечивает сопоставимость национальных статистических данных о внешней торговле на международном уровне.

3.4. Как выполнить группировку по количественным признакам?

Группировки по количественным признакам чрезвычайно разнообразны. Таковы, например, группировки населения по возрасту, среднедушевому денежному доходу, группировка рабочих по разрядам, предприятий — по численности работников, стоимости продукции и основных фондов, банков — по размеру активов и т.д. В ряде случаев группировки, на первый взгляд, казалось бы качественные, в действительности основываются на количественных признаках. Например, при группировке кредитов по срокам, на которые они предоставляются, выделяют краткосрочные, среднесрочные и долгосрочные. При отнесении в ту или иную группу руководствуются тем, что первая группа включает кредиты на срок до 1 года, вторая — от 1 года до 5 лет и третья — кредиты на срок свыше 5 лет.

Приведем еще один пример. Для характеристики трудовых ресурсов в составе постоянного населения выделяют три группы: лица

моложе трудоспособного возраста; лица трудоспособного возраста; лица старше трудоспособного возраста. Между тем основанием для формирования этих групп служит возрастной состав населения. Так, в первую группу включаются лица в возрасте до 16 лет. Во вторую группу по российскому законодательству относятся женщины от 16 до 55 лет и мужчины от 16 до 60 лет. В третью группу включаются лица пенсионного возраста: женщины старше 55 лет, мужчины старше 60 лет.

Количество выделяемых групп может определяться характером вариации изучаемого показателя. Если в качестве группировочного используется ***дискретный признак***, т.е. признак, способный принимать только некоторые определенные значения, то число выделяемых групп соответствует числу вариантов значений признака, если оно не очень велико (например, распределение рабочих предприятия по тарифным разрядам, группировка домохозяйств по размеру, распределение студентов по успеваемости (по баллам) и т.д.). В этом случае число групп определяется количеством реально существующих вариантов значений изучаемого признака. Но дискретный признак может иметь и очень большое число вариантов значений, которые не всегда могут повторяться. В таких случаях варианты значений объединяются в группы, например, при группировке предприятий по численности рабочих, по числу единиц установленного металлорежущего оборудования и т.д.

При непрерывном характере вариации группировочного признака, когда в определенных пределах признак может принимать любое значение (целое и дробное), весь диапазон изменения признака также разбивается на интервалы.

Приведем пример, когда для группировки использован непрерывный признак (табл. 3.2).

Группы кредитных организаций по величине зарегистрированного уставного капитала выделены с помощью неравных интервалов, у которых разности между верхней и нижней границей неодинаковы в разных группах. Так, во второй группе величина интервала равна 1,5 млн руб., в третьей — 3,0 млн руб., в четвертой — 5,0 млн руб., в пятой — 10 млн руб. и т.д., т.е. в приведенной группировке величина интервала постепенно увеличивается, а в последней группе верхняя граница не указывается совсем. Интервалы, в которых указана лишь одна граница (верхняя или нижняя), называются ***открытыми***. В приводимом примере открытыми являются первый (указана верхняя граница) и последний интервал (указана нижняя граница), остальные интервалы являются закрытыми, так как в них указана и верхняя, и нижняя границы. Неравные интервалы применяются при группировках, которые охватывают массу единиц неоднородной совокупности с неравномерными и

Таблица 3.2

Группировка действующих кредитных организаций по величине зарегистрированного уставного капитала по состоянию на 01.11.99 г.

Группы коммерческих банков по величине зарегистрированного уставного капитала	Количество кредитных организаций			
	На 01.01.99		На 01.11.99	
	Количество	Удельный вес к итогу (в %)	Количество	Удельный вес к итогу (в %)
До 500 тыс. руб.	90	6,1	52	3,8
От 500 тыс. руб. до 2 млн руб.	173	11,72	126	9,2
От 2 до 5 млн руб.	219	14,83	170	12,4
От 5 до 10 млн руб.	335	22,7	285	20,7
От 10 до 20 млн руб.	245	16,6	214	15,5
От 20 до 40 млн руб.	213	14,43	256	18,6
От 40 млн руб. и выше	201	13,62	272	19,8
Итого	1476	100	1375	100

Источник: Вестник банка России. 1999. № 70—71.

значительными колебаниями признака. При этом учитывается, что небольшое изменение величины группировочного признака в низших группах отражает более существенные изменения в характере группы, чем такое же по абсолютной величине различие в высших группах. Кроме того, в экономических исследованиях, как правило, с увеличением значений признака уменьшается число единиц, варианты значений признака которых находятся в интервале одной и той же длины. Поэтому применение равных интервалов может привести к тому, что число единиц в группах будет слишком мало, а закономерности в изменениях взаимосвязанных факторов могут отчетливо и не проявиться. Открытые интервалы используются обычно в тех случаях, когда признак в выделяемой группе единиц изменяется неравномерно и в широких пределах и когда отсутствуют качественные различия у отдельных единиц, включаемых в группу.

Границы интервалов проводимой с определенной целью группировки следует видоизменять в зависимости от особенностей изучаемой совокупности и временного периода, к которому относятся данные. Так, группируя предприятия по стоимости основных фондов для выделения мелких, средних и крупных предприятий, применяют интервалы в разных отраслях промышленности. Вместе с тем развитие исследуемых явлений во времени требует изменения

интервалов, т.е. предприятия, которые могли быть отнесены к крупным в изучаемой отрасли в 40-е и 50-е годы, в настоящее время скорее можно отнести к средним и даже мелким. Указанные обстоятельства находят отражение в изменении величины интервалов группировки.

Внутри типичных групп для характеристики количественных различий единиц, составляющих соответствующую группу, могут быть применены равные интервалы. В этом случае величина интервала определяется делением размаха вариации на принятое число групп. Более подробно эти вопросы рассмотрены в гл. 5.

Подытоживая вышеизложенное, можно перечислить те методологические проблемы, решение которых необходимо при практическом применении метода группировок: сначала обосновывается выбор группировочного признака или их комбинации, затем определяется число групп и величина интервалов группировки. После этого применительно к конкретной группировке устанавливается состав тех показателей, которыми должны характеризоваться выделенные группы, и, наконец, составляется макет таблицы, в которой должны быть представлены результаты группировки.

3.5. Каковы виды и функции статистических таблиц?

Статистическая таблица – форма рационального и наглядного изложения цифровых характеристик исследуемых явлений и представление ее в виде сводных статистических таблиц дает возможность характеризовать размеры, структуру и динамику изучаемых явлений. При помощи табличной формы представления данных легче осуществить сравнение и анализ цифровых показателей. Часто к статистической таблице дается общий заголовок, раскрывающий содержание таблицы, место и время, к которым относятся приводимые в таблице данные, а также единицы измерения, если они одинаковы для всех приведенных сведений. Основные элементы статистической таблицы — подлежащее и сказуемое. ***Подлежащим таблицы*** являются единицы статистической совокупности или их группы. ***Сказуемое таблицы*** отражает то, что в ней говорится о подлежащем с помощью цифровых данных. В зависимости от строения подлежащего все статистические таблицы можно разделить на три группы:

1. Таблицы простые, или перечневые, в которых содержатся сводные показатели, относящиеся к перечню единиц наблюдения, к перечню хронологических дат или территориальных подразделений.

Соответственно таблицы могут быть названы простыми перечневыми, хронологическими или территориальными.

2. Таблицы групповые, в которых статистическая совокупность разделяется на отдельные группы по какому-либо одному признаку, причем каждая из групп может быть охарактеризована рядом показателей.

3. Таблицы комбинационные, в которых совокупность разбита на группы не по одному, а по нескольким признакам.

Выбор типа таблицы всегда зависит от цели ее построения. Если таблицы используются для практических нужд планирования и управления, то в них должны содержаться сведения по тем частям, в разрезе которых ведется планирование и управление. Чаще всего этой задаче соответствуют простые таблицы, но используются также и групповые. Если же ставится задача более глубокого познания исследуемого объекта, то используются групповые и комбинационные таблицы.

В простых таблицах помещаются данные по организациям различного рода: предприятиям, стройкам, учреждениям, министерствам и т.д., имеющие, как правило, познавательное значение.

Примером простой таблицы является табл. 3.3, где приведена динамика уровня безработицы в России.

Различия в методологии расчета этих двух показателей вытекают из различий в их целевом назначении. В основе методологии, используемой службой занятости населения (уровень официальной безработицы), лежат вопросы по финансированию мероприятий, связанных с профессиональной переподготовкой и обучением, выплатой пособий по безработице и т.д. Методология МОТ нацелена на определение реальной нагрузки на рынок труда со стороны предложения рабочей силы.

В отличие от простых групповые и комбинационные таблицы обладают важными аналитическими свойствами: они позволяют

Таблица 3.3

Динамика уровня безработицы в Российской Федерации за 1994–1998 гг.

	Годы				
	1994	1995	1996	1997	1998
Уровень безработицы, рассчитанный по методологии МОТ*	7,4	8,8	9,9	11,2	13,3
Уровень официальной безработицы	2,2	3,2	3,4	2,8	2,6

* МОТ — Международная организация труда.

производить наглядные сравнения и вскрывать существенные связи и различия в развитии явлений. Таблица 3.4 является примером групповой таблицы. Ее данные свидетельствуют о том, что если до 1997 г. в России наибольшая численность населения была занята в государственном секторе, то в 1998 г. 41,9% общей численности занятых приходится на частный сектор. Идея комбинационной таблицы состоит в том, что каждую из групп в групповой таблице разбивают на подгруппы по какому-либо признаку; выделенные подгруппы могут дальше разделяться по следующему признаку и т.д. (пример комбинационной таблицы приведен в табл. 3.1).

Результаты комбинационной группировки по большому количеству признаков даже при небольшом числе интервалов становятся трудно обозримыми, но таблица делает их наглядными. Поэтому нецелесообразно составлять комбинационные таблицы по сочетанию более чем трех признаков и при количестве интервалов более четырех. Использование комбинационных таблиц и системы взаимосвязанных группировок позволяет провести глубокий и всесторонний анализ сложных общественных явлений.

Группировка, осуществляемая не последовательно по отдельным признакам, как при комбинационной группировке, а одновременно по комплексу признаков, называется многомерной. Как уже сказано, характеристика одной и той же качественной стороны изучаемого явления может быть дана с помощью набора признаков. Например, для характеристики технического уровня развития предприятий могут быть использованы следующие показатели:

Таблица 3.4

Распределение занятого населения России по секторам экономики

	Годы			
	Млн человек		В % к итогу	
	1997	1998	1997	1998
Занято в экономике, всего	64,6	63,6	100	100
В том числе по секторам				
Государственные и муниципальные предприятия и организации	25,9	24,4	40,1	38,3
Частный сектор	25,8	26,6	39.9	41,9
Общественные организации, фонды	0,4	0,4	0,6	0,6
Совместные предприятия и др.	12,5	12,2	19,4	19,2

Источник: Российский статистический ежегодник. 1998. С. 173.

удельный вес активной части промышленно-производственных основных фондов, удельный вес автоматических машин и оборудования в составе рабочих машин и оборудования, электровооруженность труда, машиновооруженность рабочих, степень охвата механизированным трудом, коэффициент обновления машин и оборудования и т.д.

Таким образом, характеризуя каждую единицу совокупности (предприятие) набором, признаков, можно рассматривать эту единицу как точку в m-мерном пространстве, при этом задачей многомерной группировки будет выделение точек, составляющих однородные группы единиц. Мерой сходства единиц могут служить различные критерии. В зависимости от выбранного критерия существуют различные методы многомерной группировки.

Применение методов многомерной группировки связано с большой вычислительной работой и требует использования электронной вычислительной техники. С помощью специальных алгоритмов на ЭВМ осуществляется формирование групп, в которых единицы совокупности объединяются на основании близости по всему комплексу признаков. В табл. 3.5 приведены результаты группировки предприятий отрасли по уровню технического развития и производительности труда.

Таблица 3.5

Распределение предприятий по уровням технической оснащенности и производительности труда

Группы предприятий по уровню технической оснащенности	Группы предприятий по уровню производительности труда			Итого
	Ниже среднего по отрасли уровня	Среднего по отрасли уровня	Выше среднего по отрасли уровня	
Ниже среднего по отрасли уровня	9	8	8	25
Среднего по отрасли уровня	5	6	1	12
Выше среднего по отрасли уровня	6	3	7	16
Итого	20	17	16	53

Примечание: Выделение однородных по техническому уровню групп предприятий было осуществлено с помощью метода кластерного анализа по восьми показателям технического уровня развития.

Анализ данных таблицы позволяет выделить группы предприятий, добившихся наибольшего эффекта в своей деятельности и группы предприятий, располагающих резервами роста производительности труда за счет лучшего использования технического потенциала. Это прежде всего те шесть предприятий, которые имеют техническую оснащенность выше средней по отрасли, но уровень производительности труда здесь ниже среднего по отрасли. В то же время восемь предприятий с уровнем технической оснащенности ниже среднеотраслевого имеют уровень производительности труда выше среднеотраслевого, что позволяет говорить об их высокоэффективной деятельности.

Такие группировки дают возможность, безусловно, лишь в общем оценить результаты деятельности предприятий соответствующих групп. Однако они делают очевидной необходимость дифференцированного подхода к оценке результатов деятельности предприятий отрасли, располагающих примерно одинаковым техническим потенциалом, но различающихся уровнем эффективности использования труда.

3.6. Какое значение имеют графики в статистике и каковы общие правила их построения?

Графики являются одной из форм представления статистической информации. Они находят широкое применение при обобщении и анализе данных, так как позволяют наглядно и доступно отразить результаты обработки большого объема информации.

Статистический график представляет собой условное изображение, при помощи которого с помощью геометрических образов дается характеристика определенных показателей. Иллюстрация данных в виде графиков позволяет облегчить восприятие статистической информации и способствует правильному ее толкованию.

Значение графического метода при обобщении и анализе статистических данных заключается в установлении закономерностей развития социально-экономических явлений, выявлении взаимосвязей признаков, анализе тенденций изменения явления во времени, оценке выполнения плана. Графики также используются для характеристики структуры явлений и их пространственных сопоставлений. Следовательно, графики являются незаменимым средством для научного обобщения статистической информации и ее популяризации.

Для графического изображения статистических данных используются различные виды графиков, классификация которых представлена на рис. 3.1. В соответствии с этой классификацией графики

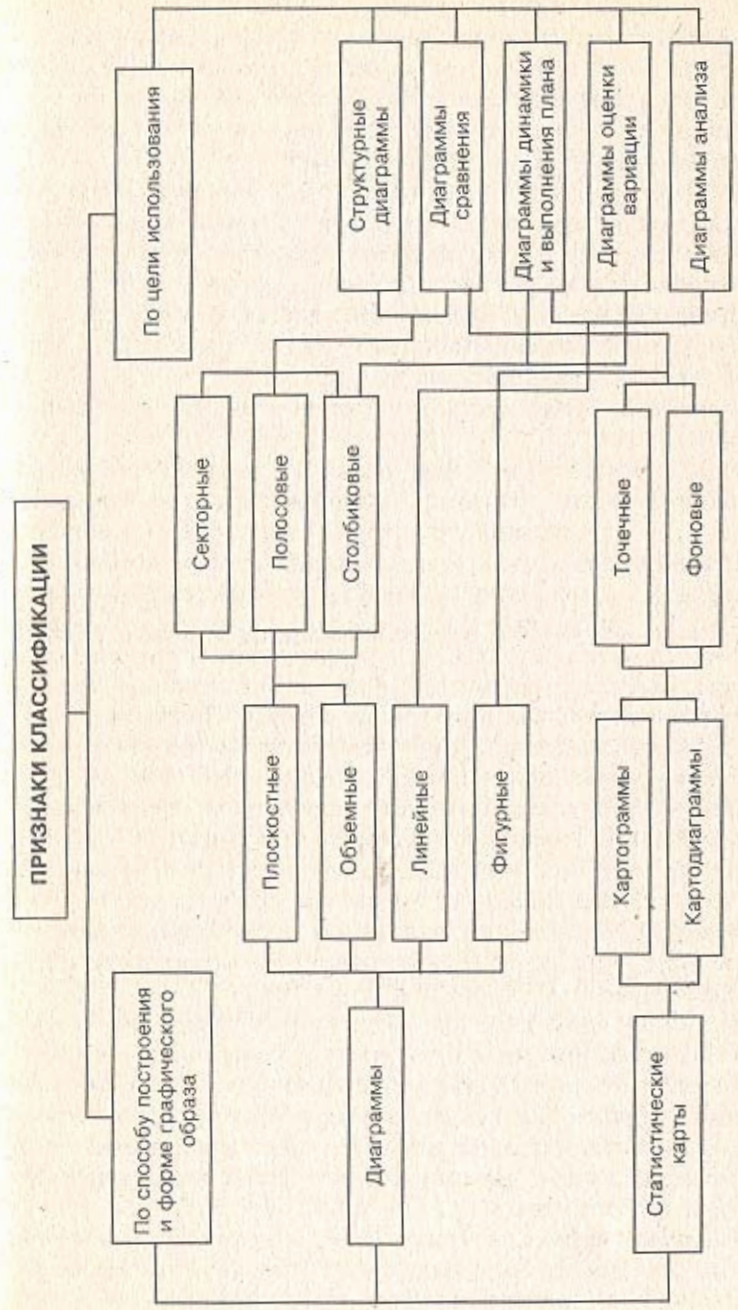


Рис. 3.1. Классификация графиков

подразделяются по следующим признакам: цели использования, способу построения и форме графического образа.

По ***способу построения*** графики делятся на диаграммы и статистические карты. В зависимости от ***целей использования*** выделяют структурные диаграммы, диаграммы сравнения, динамики и выполнения плана. Кроме того, диаграммы используются для оценки вариации, выявления и анализа взаимосвязей.

Для обеспечения практического использования графиков необходимо соблюдать следующие правила их построения: выбрать графический образ; задать поле графика, пространственные и масштабные ориентиры; ввести экспликацию графика.

Графический образ, т.е. вид графика, представляет собой те точки, линии, геометрические фигуры и т.д., которые используются для изображения статистических показателей.

Поле графика — плоскость, на которой изображены графические образы.

Пространственные ориентиры задаются в виде системы координат. Наиболее распространенной является прямоугольная система координат. Для изображения периодических (сезонных и цикличесских) колебаний используются радиальные диаграммы, которые строятся в полярных координатах. Пример радиальной диаграммы приведен на рис. 3.2. Время (в рассматриваемом примере 1—12 месяцев) отсчитывается по часовой стрелке по окружности, а уровень показателя откладывается от центра круга (точки 0) в заданном масштабе по соответствующему каждому месяцу радиусу. Построенная радиальная диаграмма наглядно изображает наличие сезонных колебаний анализируемого признака. Кроме того, нанесение на график значений признака за несколько лет позволяет одновременно судить и о тенденции его изменения (росте или снижении).

Масштабные ориентиры зависят от масштаба и масштабной шкалы графика. ***Масштаб*** — это мера перевода числовой величины в графическую. Масштабной шкалой является линия, на которой в определенном порядке нанесены штрихи и соответствующие им числа, расположенные строго под штрихами.

Масштабные шкалы бывают обычными и логарифмическими. Так, для изображения временных рядов со значительными различиями в величине уровней строится полулогарифмический график, у которого одна из шкал логарифмическая. На этом графике по оси абсцисс в обычном масштабе откладываются периоды или моменты времени, а по оси ординат в логарифмическом масштабе — уровни временного ряда.

Экспликация графика — это словесное описание его содержания. Оно включает название рисунка, подписи вдоль масштабных шкал и пояснения к отдельным компонентам графика.

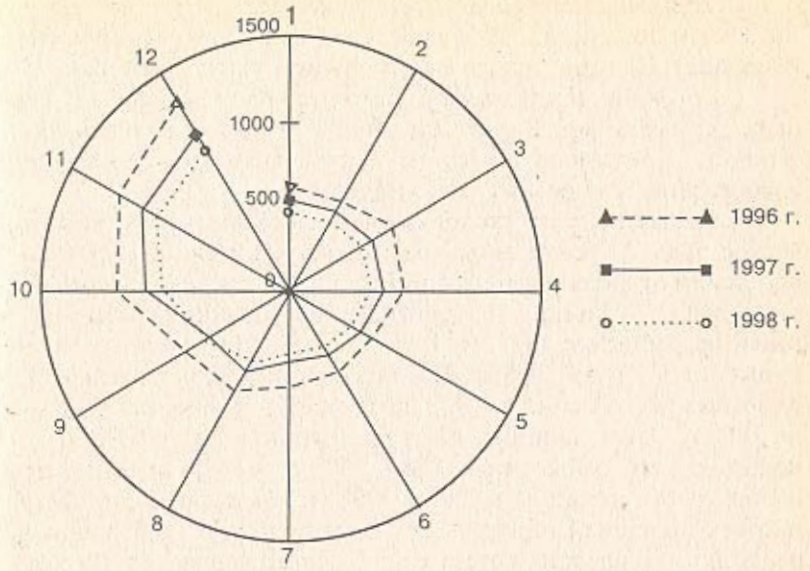


Рис. 3.2. Производство продуктов животноводства по месяцам года (в тыс.тонн) за 1996—1998 гг.

Источник: Вопросы статистики. 1999. № 5.

Получившие широкое распространение в современных условиях пакеты прикладных программ компьютерной графики значительно облегчают задачи исследователя при построении и практическом применении графиков. Наиболее известными пакетами прикладных программ являются: *SPSS, STATISTIKA, STADIA, STATGRAPHICS, Excel, Harvard graphics* и др.

3.7. Какие виды графиков получили наибольшее распространение?

Рассмотрение видов графиков начнем с диаграмм, как наиболее простых видов графиков и получивших широкое распространение в экономико-статистическом анализе. Диаграммы применяются для наглядного сопоставления социально-экономических явлений в пространстве (по территориям, странам, фирмам, различным видам продукции), а также для анализа их динамики.

Чаще других для этих целей на практике применяются столбиковые диаграммы. Каждый столбик соответствует по величине уров-

ню исследуемого статистического показателя, что позволяет сравнивать эти показатели. Столбики строятся в прямоугольной системе координат. На горизонтали находится основание столбиков, ширина и расстояние между ними выбираются произвольно, но должны быть одинаковыми. Высота столбиков меняется в зависимости от величины показателя. На одном графике возможно одновременное изображение нескольких показателей.

В качестве примера столбиковой диаграммы рассмотрим графическое представление динамики оттока капитала и валютных поступлений от внешней торговли России. Из графика, изображенного на рис. 3.3, видно, что валютные поступления от внешней торговли возрастали с 1993 по 1996 г., затем имела место тенденция снижения в 1997 и 1998 гг. Наблюдающийся незначительный рост валютных поступлений в 1999 г. не позволил однако достигнуть уровня 1996 г. Отток капитала имел тенденцию к росту до 1997 г. При некотором его снижении в 1998 и 1999 гг. можно отметить превышение оттока капитала в 1997 и 1998 гг. над валютными поступлениями от внешней торговли. Это свидетельствует о больших потерях валютных средств, которые необходимо направлять на развитие производства и социальной сферы в России.

Отличие полосовых диаграмм от других состоит в том, что масштабная шкала располагается по горизонтали и определяет величину полос по длине. При построении столбиковых и полосовых диаграмм необходимо, чтобы шкала для определения размера столбика (полосы) была непрерывной. Область применения этих видов диаграмм одинакова.



Рис. 3.3 Сравнение оттока капитала и валютных поступлений от внешней торговли Российской Федерации

Источник: James Fenkner, *Maket Weekly//Тройка–диалог*. 1999. 4.

Для целей сравнения также используются диаграммы в виде различных геометрических фигур, по площади которых устанавливается размер социально-экономического явления. Наиболее распространенными являются квадратные и круговые диаграммы.

Для графического изображения структуры социально-экономических явлений используются секторные, столбиковые и полосовые диаграммы. Секторная диаграмма является одним из наиболее распространенных видов структурных диаграмм. Анализ структуры осуществляется в результате сопоставления различных частей целого при помощи площадей, образуемых секторами круга.

Круговая секторная диаграмма строится путем деления круга на секторы пропорционально удельному весу частей в целом. Сумма удельных весов равна 100%, что соответствует общему объему анализируемого явления и составляет 360°. Поэтому размер каждого сектора определяется величиной угла из расчета, что 1% соответствует 3,6°. Для более четкого изображения секторов рекомендуется использовать штриховку. Примером секторной диаграммы является представленная на рис. 3.4 диаграмма структуры денежных доходов населения Российской Федерации в первом квартале 1997 г. и первом квартале 1998 г. Из диаграмм видно, что если в первом квартале 1997 г. наибольший удельный вес в общем объеме денежных доходов составляли другие доходы (43,7%), то уже в первом квар-



Рис. 3.4. Структура денежных доходов населения Российской Федерации в первом квартале 1997 г. (а) и первом квартале 1998 г. (б)

Источник: Текущие тенденции в денежно-кредитной сфере. Статистико-аналитические материалы ЦБ РФ. № 4 (62). М., 1998.

тале 1998 г. – оплата труда (47,2%). Удельный вес пенсий, пособий, стипендий невелик. В то же время наблюдается рост их удельного веса в первом квартале 1998 г. по сравнению с первым кварталом 1997 г. (с 14,8 до 16,0% соответственно).

Для анализа динамики социально-экономических явлений и процессов строятся линейные, столбиковые, ленточные, круговые, радиальные и другие диаграммы. В частности, для рядов динамики с неравноотстоящими уровнями можно воспользоваться столбиковыми или круговыми диаграммами. В случае большого числа равноотстоящих уровней более удобными для восприятия являются линейные диаграммы, так как они позволяют оценить тенденцию или колебания в развитии явления. Кроме того, используя темпы роста, на одном графике можно сравнивать динамику разных показателей.

Для построения линейных диаграмм в прямоугольной системе координат по оси абсцисс откладываются отрезки, соответствующие датам или периодам времени, по оси ординат – уровни ряда или темпы роста. По обозначенным точкам строят ломаную линию. На рис. 3.5 показаны две ломаные линии, характеризующие динамику объемов промышленной продукции и железнодорожных перевозок грузов в Российской Федерации за период с января 1995 по октябрь 1999 г. Для анализа были использованы базисные темпы роста к январю 1995 г.

Из графика видно, что начиная с октября 1995 г. и по сентябрь 1999 г. объем перевозок грузов снижался по отношению к январю 1995 г. В динамике объема производства промышленной продукции наблюдаются существенные колебания, среди которых выделяется незначительный рост в марте и октябре каждого года. В то же время с октября 1995 г. по сентябрь 1999 г. темпы роста объема промышленной продукции превышали темпы роста объема перевозок грузов. При сравнении динамики этих показателей можно отметить совпадение тенденций их изменения, что наиболее очевидно в период с апреля по сентябрь 1999 г.

В отдельных случаях изображение на одном графике двух кривых позволяет оценить динамику третьего показателя, как разность первых двух. Так, например, при изображении кривых динамики экспорта и импорта площадь между ними будет соответствовать величине сальдо внешней торговли.

Если на одном графике сравнивается динамика двух показателей, имеющих разные единицы измерения, то масштабные шкалы строятся слева и справа.

Статистические карты используются для характеристики распределения явления на определенной территории. К статистическим картам относятся картограммы и картодиаграммы.

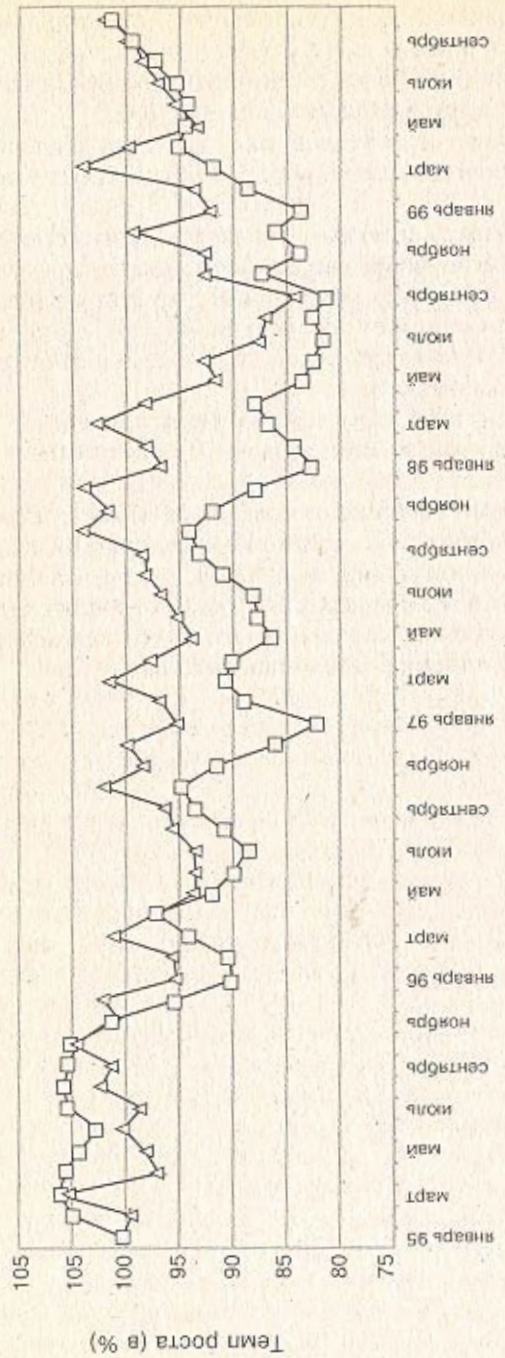


Рис. 3.5 Динамика объема промышленной продукции и объема железнодорожных перевозок грузов в Российской Федерации (в % к январю 1995 г.)

Источник: Россия 1999. Экономическая конъюнктура. М.: Центр экономической конъюнктуры при Правительстве Российской Федерации.

Картограмма представляет собой географическую карту, на которой изображается интенсивность размещения социально-экономического явления в пределах каждой единицы территориального деления. Картограммы делятся на фоновые и точечные.

Фоновые картограммы штриховкой или окраской различной яркости показывают степень интенсивности или плотности показателя по территории.

Особенностью ***точечных картограмм*** является то, что величина анализируемого показателя изображается с помощью точек. Точка соответствует определенному значению, принятому для характеристики единицы совокупности, или определенному количеству единиц. Таким образом, можно получить представление о плотности анализируемого признака по регионам.

Фоновые картограммы используются для характеристики средних или относительных показателей, точечные — для объемных показателей.

Картодиаграмма — это сочетание географической карты с диаграммой. В качестве изобразительных знаков в картодиаграммах используются фигурные диаграммы (круги, столбики, квадраты и т.д.), размещаемые на контуре географической карты. С помощью картодиаграмм можно отразить более сложные статистико-географические сопоставления по сравнению с картограммами.

Глава 4
АБСОЛЮТНЫЕ, ОТНОСИТЕЛЬНЫЕ И СРЕДНИЕ ВЕЛИЧИНЫ

4.1. Какова роль статистических показателей в управлении экономикой и в чем состоят общие принципы их построения?

Специфика статистического изучения действительности состоит в том, что при конструировании показателей, отражающих количественную сторону массовых общественных явлений, статистика опирается на сформулированные экономической наукой понятия, отражающие сущность явлений. Процесс перехода от понятий и категорий экономической науки к системе экономико-статистических показателей представляет собой сложный взаимосвязанный процесс теоретического и эмпирического познания. Понятия экономической науки являются определяющими в этом процессе, ибо лишь после того как выяснена сущность и отличительные особенности того или иного явления, имеет смысл иллюстрировать его развитие посредством обработанных надлежащим образом статистических данных.

Экономико-статистические показатели содержат количественную характеристику тех или иных свойств экономических явлений. В отличие от понятий их следует конкретизировать применительно к тем реальным явлениям, которые отражаются этим понятием. Таким образом, понятие надо свести к простым и однородным сторонам, свойствам и признакам, которые допускают операции сравнения или измерения.

Показатель формализует содержание изучаемых сторон социально-экономических явлений и представляет собой модель их количественной характеристики. Чем сложнее исследуемое явление, тем труднее оно поддается формализации и моделированию. С помощью показателей определяется, что, где, когда и каким образом следует численно измерить. Каждый статистический показатель с возможной точностью должен соответствовать сущности того явления, которое должно быть измерено с его помощью. Например, измерение объема продукции промышленности требует предварительного установления тех видов деятельности промышленного предприятия, которые будут учтены в составе промышленной про-

дукции, и определения тех результатов этой деятельности, которые могут быть включены в ее объем.

Необходимость рассмотрения исследуемого объекта во всех его связях и отношениях приводит к тому, что для получения целостной статистической характеристики изучаемых явлений применяются ***системы статистических показателей***. Важнейшая особенность системы показателей — их содержательное единство, связанное с характеристиками единого объекта исследования. Так, система стоимостных показателей продукции промышленного предприятия включает следующие показатели: товарная продукция, отгруженная продукция, реализованная продукция, чистая продукция, добавленная стоимость и др. Совершенствование управления экономикой обусловило переход от преимущественной ориентации на показатель валовой продукции к показателям, оценивающим вклад непосредственно данного предприятия в производство продукции, таким, как добавленная стоимость или чистая продукция.

С одной стороны, переход к количественному описанию явлений действительности обогащает и уточняет понятие о явлениях. С другой — при построении показателей почти всегда приходится дополнительно упрощать, схематизировать реальные явления, а потому статистические показатели лишь с известной степенью приближения отражают объективную реальность. Все это делает очень важной постоянную работу по совершенствованию системы показателей. Развитие систем статистических показателей происходит в соответствии с развитием отражаемой объективной действительности и в результате углубления процесса познания реальных систем.

При формировании систем статистических показателей должны быть четко сформулированы значение и область применения соответствующих показателей, определены функции, которые им надлежит выполнять.

Статистические показатели — это величины, адекватно характеризующие с разных сторон отображаемое явление в конкретных условиях времени и места. В этом определении прежде всего показана учетная функция статистических показателей, реализация которой связана с отражением в них объективных свойств анализируемых явлений. Изучая отчетные данные по действующей системе показателей, руководители разного уровня управления должны получить объективную информацию о реальном состоянии и тенденциях развития организаций, отраслей и экономики в целом. Планирование деятельности предприятий, фирм, корпораций опирается на действующую систему показателей. ***Плановые показатели,*** осуществляя директивную функцию, призваны ориентировать руководителей и работников предприятий на выполнение поставленных задач. Составы систем ***плановых и учетно-статистических***

показателей различаются и довольно значительно, что связано с необходимостью проведения комплексного и всестороннего анализа деятельности предприятий. С помощью анализа выявляются неиспользованные резервы, определяются пути устранения нежелательных отклонений и т.д. При этом нужно иметь в виду, что отчетные показатели должны быть сопоставимы с плановыми по методологии их определения.

Статистические показатели призваны способствовать усилению воздействия плановых показателей на деятельность производственных коллективов, т.е. в определенной степени выполнять также стимулирующую функцию. Таким образом, система экономико-статистических показателей в управлении предприятиями призвана выполнять три основные функции: ***директивную, учетную и стимулирующую***. В практике хозяйственной деятельности некоторые из применяемых показателей выполняют одновременно все названные функции. Вместе с тем необходимо отметить, что статистические показатели не только играют важную роль в формировании информационного обеспечения управления на разных уровнях, но и используются для повышения информированности всего населения страны о процессах, происходящих в экономическом и социальном развитии общества. С этих позиций очень важна ***познавательная функция*** статистических показателей.

4.2. Какие группы обобщающих статистических показателей выделяют?

В процессе статистического наблюдения получают данные о значениях тех или иных признаков, характеризующих каждую единицу исследуемой совокупности. Для характеристики совокупности в целом или ее частей данные по отдельным единицам совокупности подвергают сводке и в результате получают обобщающие показатели, в которых отражаются результаты познания количественной стороны изучаемых явлений.

Обобщающие показатели могут быть представлены ***абсолютными, относительными*** и ***средними величинами***. Они широко используются в планировании и анализе деятельности предприятий и фирм, отраслей и экономики в целом. Многообразная характеристика всех сторон исследуемых социально-экономических явлений может быть дана лишь с помощью всех видов обобщающих показателей. Вместе с тем каждый вид показателей имеет определенное значение и занимает различное место в процессе познания реальной действительности.

Путем непосредственного суммирования первичных данных получают обобщающие ***абсолютные показатели***, которые характеризуют численность совокупности и объем (размер) изучаемого явления в конкретных границах времени и места. Например, численность населения России на начало 1999 г. составила 146,3 млн человек, валовой внутренний продукт (ВВП) в фактически действующих ценах был равен в 1998 г. 2685 трлн руб. Первый из приведенных показателей характеризует численность населения по состоянию на определенный момент; получают такие абсолютные величины по результатам сводки данных единовременного наблюдения. Второй из приведенных показателей характеризует размер ВВП за определенный период; он получен как результат сводки данных текущего наблюдения.

Абсолютные показатели являются всегда именованными числами, т.е. имеют какую-либо единицу измерения. ***Натуральные единицы измерения*** применяют в тех случаях, когда единица измерения соответствует потребительским свойствам продукта. Например, производство цемента оценивается в тоннах, тканей — в квадратных метрах, автомобилей — в штуках и т.д. Натуральные единицы могут быть и составными (сложными). Например, отработанное рабочими время учитывается в человеко-днях и человеко-часах, грузооборот автомобильного и железнодорожного транспорта — в тонно-километрах и т.п. Составные единицы отражают сочетание двух различных сторон явления. Так, при учете затрат рабочего времени отражается совместное измерение численности рабочих и времени их работы, при определении объема работы транспорта — измерение объемов грузов и протяженности перевозок.

При учете продукции в натуральном выражении нередки случаи, когда применяются различные единицы измерения для одного и того же вида продукции. Это делается для того, чтобы полнее охарактеризовать потребительское назначение продукции и изменение ее состава. Например, производство электродвигателей учитывается в штуках и киловаттах мощности, бумаги — в тоннах и квадратных метрах, стальных труб — в тоннах и погонных метрах и т.п.

Если некоторые разновидности продукции обладают общностью основного потребительского свойства, обобщенные итоги по выпуску этих разновидностей продукции можно получить, используя ***условно-натуральные единицы***. В этом случае одна из разновидностей принимается в качестве единого измерителя, а другие приводятся к этому измерителю с помощью соответствующих коэффициентов пересчета. Например, в тоннах условного топлива определяется общий объем потребленного на предприятии топлива, при определении объема производства минеральных удобрений пересчет производится на стандартное или 100%-ное содержание питательного вещества и т.п.

При обобщении учетных данных даже на уровне предприятий, а тем более на уровне отраслей и народного хозяйства широко используются ***стоимостные (денежные) единицы измерения***. Для получения общего объема продукции в денежном выражении количество единиц каждого вида продукции в натуральном выражении умножается на цену соответствующего вида, а затем полученные произведения суммируют во всем видам. При определении стоимостных показателей объема продукции абсолютные величины получают расчетным путем. Особенно это касается определения таких обобщающих показателей, как чистая продукция промышленности, прибыль, объем валового национального продукта и др. Так, объем чистой продукции определяют путем вычитания из стоимости товарной продукции стоимости материальных затрат на ее производство. Прибыль от реализации продукции промышленного предприятия получают как разность между выручкой от реализации продукции по оптовым ценам предприятий и ее полной себестоимостью.

Расчетный метод определения абсолютных величин может опираться также на балансовый метод. Например, наличие оборудования на конец года определяется путем прибавления к количеству оборудования на начало года числа единиц введенного оборудования и вычитания числа единиц выбывшего оборудования. Таким образом, абсолютные величины получают непосредственным подсчетом по данным статистического наблюдения или расчетным путем. ***Абсолютные статистические показатели*** могут быть измерены с различной степенью точности. С переходом к более высоким ступеням обобщения применяются и более укрупненные единицы измерения. Например, производство металлорежущих станков на промышленном предприятии учитывается в штуках, а в масштабах страны — в миллионах штук; государственные закупки зерна в хозяйствах измеряются в тоннах, а по России в целом — в тысячах тонн.

Использование одноименных единиц измерения исследуемых показателей является важным исходным условием при международных сравнениях. Вместе с тем одноименные единицы измерения объема производства однотипной продукции в разных странах могут быть неодинаковыми по величине или методологии учета. Это касается, например, показателей производства электроэнергии, выпуска строительного кирпича и ряда других продуктов.

Хотя абсолютные величины играют важную роль в практической и познавательной деятельности человека, анализ фактов обязательно приводит к необходимости различного рода сопоставлений. И тогда абсолютные показатели, характеризующие те или иные

изучаемые явления, рассматриваются не только самостоятельно, но и в сравнении с другим показателем, который принимается за масштаб оценки, или базу сравнения.

Сопоставление статистических данных осуществляется в различных формах и по разным направлениям. В соответствии с задачами и направлениями сопоставления статистических данных применяются различные виды ***относительных величин***.

Причем сопоставлять можно одноименные показатели, относящиеся к различным периодам, различным объектам или разным территориям. Результат такого сопоставления может быть представлен коэффициентом (база сравнения принята за единицу) или выражен в процентах и показывает, во сколько раз или на сколько процентов сравниваемый показатель больше или меньше базисного.

4.3. Каково значение разных видов относительных величин?

В результате сопоставления одноименных показателей получают следующие относительные величины: относительные величины динамики; относительные величины выполнения плана и планового задания; относительные величины структуры; относительные величины координации; относительные величины наглядности; относительные величины интенсивности.

Относительная величина динамики характеризует изменение показателя во времени и показывает, во сколько раз увеличился (или уменьшился) уровень показателя по сравнению с каким-либо предшествующим периодом. Для расчета относительной величины динамики определяют отношение уровней, характеризующих изучаемое явление в разные периоды времени. Как правило, при анализе опираются на данные за ряд периодов. Так, в статистической отчетности содержатся показатели за каждый месяц отчетного года или за соответствующий квартал. На основе обобщения квартальных и месячных данных получаем итоги за каждый год. Статистические показатели за каждый квартал отчетного года можно сравнить или с уровнем непосредственно предшествующего года, или с уровнем соответствующего квартала предшествующего года, или же с уровнем четвертого квартала предшествующего года, т.е. рассчитываются относительные величины с переменной или постоянной базой сравнения.

Относительные величины имеют форму коэффициентов, если они исчисляются путем деления сравниваемой величины на базу сравнения. Если коэффициент умножить на 100, то получим результат сопоставления в процентах. Таковы относительные величины, характеризующие динамику внешнеторгового оборота (см. графы 5 и 6 табл. 4.1). Рассчитанная относительная величина динамики за 1998 г. показывает, например, что по сравнению с 1997 г. оборот внешней торговли России снизился на 17,66%. Эта величина исчисляется как разность между величиной темпа роста (коэффициент роста, выраженный в процентах) и 100% и называется относительным приростом (в приводимом примере — это относительное снижение (82,34% — 100% = −17,66%)). В абсолютном выражении снижение внешнеторгового оборота составило 28,6 млрд долл. США. Данные графы 6 табл. 4.1 свидетельствуют, что если в 1997 г. имела место тенденция роста объема внешнеторгового оборота по кварталам года, то в 1998 г. наблюдалось устойчивое снижение этого показателя на протяжении всех четырех кварталов. Наибольшее снижение (на 21,75%) имело место в I квартале 1998 г. Вместе с тем в сравнении с I кварталом 1997 г. внешнеторговый оборот не изменился (см. графу 5 табл. 4.1). Это может быть связано с наличием так называемых сезонных колебаний, что требует анализа динамики внешнеторгового оборота за более длительный временной интервал (о сезонных колебаниях подробнее см. в гл. 8).

В планах предприятий задания устанавливаются как в абсолютных показателях, так и в виде относительных величин, которые показывают, во сколько раз или на сколько процентов должна увеличиться (уменьшиться) величина показателя по плану в сравнении с его уровнем в предшествующем периоде. Например, по плану темп роста экспорта продукции предприятия должен был составить в 1999 г. 106,1% (в процентах к предшествующему году), относительная величина динамики объема экспорта продукции в 1999 г. составила 107,3%. Сравнение двух относительных величин — планового задания и динамики — свидетельствует о перевыполнении плана по экспорту продукции предприятия в 1999 г. Степень выполнения плана оценивается с помощью ***относительной величины выполнения плана***, которую получают путем деления фактического уровня показателя в отчетном периоде на его уровень, запланированный на этот же период.

Рассмотрим, как связаны между собой относительные величины планового задания, выполнения плана и динамики.

Обозначим как y^0 фактический уровень показателя в базисном периоде; как $y^{пл}$ планируемый уровень показателя на отчетный период; как y^1 фактический уровень показателя в отчетном периоде. Тогда в принятых обозначениях относительные величины могут быть представлены следующими соотношениями:

Таблица 4.1

Внешнеторговый оборот России*

Год и квартал	Внешнеторговый оборот, всего (в млрд долл. США)	В том числе		Коэффициент роста внешнеторгового оборота в процентах		Удельный вес, (в %)	
		Экспорт	Импорт	к соответствующему периоду предшествующего года	к предыдущему периоду	Экспорта	Импорта
1	2	3	4	5	6	7	8
1997 г.							
I кв.	36,7	21,1	15,6	—	—	57,49	42,51
II кв.	37,9	20,4	17,5	—	103,27	53,83	46,17
III кв.	40,4	21,6	18,8	—	106,60	53,47	46,53
IV кв.	46,9	25,1	21,8	—	116,09	53,52	46,48
Итого за год	161,9	88,2	73,7	—	—	54,48	45,52
1998 г.							
I кв.	36,7	18,4	18,3	100,00	78,25	50,14	49,86
II кв.	36,4	18,7	17,7	96,04	99,18	51,37	48,63
III кв.	31,5	17,8	13,7	77,97	86,54	56,51	43,49
IV кв.	28,7	19,3	9,4	61,19	91,11	67,25	32,75
Итого за год	133,3	74,2	59,1		82,34	55,66	44,34

* На основе данных таможенной статистики и с учетом официально нерегистрируемого экспорта (импорта)//Вопросы статистики. 1999. №11. С. 71.

Относительная величина планового задания = $y^{пл}/y^{0}$;

Относительная величина выполнения плана = $y^{1}/y^{пл}$;

Относительная величина динамики = y^{1}/y^{0}.

Следовательно, относительная величина динамики может быть получена путем умножения относительных величин планового задания и выполнения плана:

$$y^{1}/y^{0} = (y^{пл}/y^{0})(y^{1}/y^{пл}).$$

Таким образом, в приводимом примере на основе двух известных относительных величин: динамики и планового задания — может быть определена относительная величина выполнения плана по росту экспорта продукции предприятия:

$$y^{1}/y^{0} = (y^{1}/y^{0})/(y^{пл}/y^{0});\ y^{1}/y^{пл} = 1{,}073 : 1{,}061 = 1{,}0113,$$

т.е. относительная величина выполнения плана составила 101,13%, а это означает, что план по росту экспорта продукции предприятия был перевыполнен в 1999 г. на 1,13%.

Относительные величины структуры характеризуют долю отдельных частей в общем объеме совокупности, их рассчитывают как отношение числа единиц (или объема признака) в отдельных частях совокупности к общей численности единиц (или объему признака) по всей совокупности. Относительные величины структуры рассчитываются по сгруппированным данным. Например, по результатам переписи определяют долю мужчин и женщин, городского и сельского населения и т.д.:

$$d_i = \frac{y_i}{\sum_{i=1}^{K} y_i},$$

где d_i — доля i-й группы;

y_i — объем признака (число единиц) i-й группы;

K — число групп.

Расчет относительных величин структуры за несколько периодов позволяет выявить структурные сдвиги. В графах 7 и 8 табл. 4.1 приведен расчет относительных величин структуры. В них показана доля экспорта и импорта в объеме внешнеторгового оборота России за период с I квартала 1997 г. по IV квартал 1998 г.

Если в 1997 г. доля экспорта во II, III и IV кварталах оставалась достаточно устойчивой, то в 1998 г. в первых кварталах наблюдалось даже снижение доли экспорта по сравнению с 1997 г., а затем, особенно в IV квартале, произошло резкое увеличение доли экспорта до 67,25%. Это обстоятельство несомненно связано с резким изменением (почти в три раза) курса доллара после 17 августа 1998 г.

В целях анализа наряду с соотношением части и целого (показателями структуры) определяют соотношение двух частей одного целого. Относительные величины, характеризующие соотношение частей одного целого, называют ***относительными величинами координации***. К таким величинам относятся, например, показатели, характеризующие соотношение численности городского и сельского населения, численности рабочих и служащих, численности мужчин и женщин, величин заемного и собственного капитала банка, показатели демографической нагрузки и т.д. Относительные величины координации нередко характеризуются числом единиц одной части на 100 или 1000 единиц другой части. Например, в 1997 г. на 1000 руб. экспорта приходилось 836 руб.

((73,7 × 1000)/88,2), а в 1998 г. на каждые 1000 руб. экспорта приходилось уже 797 руб. импорта ((59,1 × 1000)/74,2).

Относительные величины наглядности отражают результаты сопоставления одноименных показателей, относящихся к одному и тому же периоду (или моменту) времени, но к разным объектам или территориям. Этот вид относительных величин применяется для сравнительной оценки уровня развития стран и регионов, а также при оценке результатов деятельности отдельных предприятий отрасли. Обычно их исчисляют в процентах или кратных отношениях, показывающих, во сколько раз одна из сравниваемых величин больше (или меньше) другой.

Данные табл. 4.2 свидетельствуют об увеличении разрыва в потреблении в России и США всех приведенных в таблице продуктов питания. Особенно резко отличается потребление мяса и фруктов (соответственно 0,28 и 0,19 от уровня США в 1998 г.). В то же время значительно выше уровень потребления в России по сравнению с США таких продуктов, как хлеб и картофель (соответственно 1,35 и 2,34 от уровня США в 1998 г.). Это свидетельствует о низком качестве потребления продуктов питания в России по сравнению с потреблением в США и имеющейся тенденции снижения этого качества.

Относительные величины наглядности находят широкое применение не только при международных сопоставлениях, но и при сравнительной оценке показателей деятельности предприятий и различных регионов. Так, могут сравниваться результаты работы

Таблица 4.2

Потребление некоторых продуктов питания на душу населения (в кг)

Год	Страна	Продукты питания				
		Мясо	Картофель	Овощи	Фрукты	Хлеб
1985	Россия	62	109	98	40	119
	США	112	56	127	95	98
1998	Россия	35	145	60	20	140
	США	125	62	118	107	104
Отношение уровня потребления продуктов питания в России к уровню потребления в США						
1985		0,55	1,95	0,77	0,42	1,21
1998		0,28	2,34	0,51	0,19	1,35

предприятий различных форм собственности. Безусловный интерес представляет сопоставление цен на продукцию государственных и частных предприятий, средней заработной платы работников этих предприятий и т.д. При этом за базу сравнения принимают цену и другие показатели государственных предприятий.

Отношения между разноименными абсолютными величинами называют ***относительными величинами интенсивности***. В их числе можно назвать показатели уровня жизни населения, к которым относятся приведенные в табл. 4.2 показатели потребления продуктов питания и непродовольственных товаров на душу населения; показатели обеспеченности населения предметами культурно-бытового и хозяйственного назначения длительного пользования в расчете на 100 семей или 1000 человек населения; показатели обеспеченности населения жильем и т.д. Учитывая экономическую сущность относительных величин интенсивности, их можно было бы назвать показателями уровня экономического и социального развития.

В отличие от относительных величин, являющихся результатом сопоставления одноименных показателей и представляемых с помощью коэффициентов или процентов, относительные величины интенсивности являются именованными числами (число аварий на 10 тыс. человек населения, производство хлопчатобумажных тканей в квадратных метрах на душу населения и т.д.).

Важно отметить, что в процессе экономико-статистического анализа абсолютные и относительные величины должны рассматриваться во взаимосвязи, т.е. пользоваться относительными величинами нужно не формально, а представлять при этом, какая абсолютная величина скрывается за каждым относительным показателем. Особенно важно соблюдать это положение при расчете относительных величин динамики.

Смена базы сравнения при расчете относительных величин может иногда ввести в заблуждение неискушенного читателя или просто скрывать действительное положение вещей. Например, в связи с финансовыми затруднениями фирма снизила заработную плату менеджеру, получавшему 400 долл. США на 10%. На следующий год его заработная плата повысилась на 10%. Означает ли это, что менеджер будет получать вновь 400 долл.? Ответ на вопрос, безусловно, отрицательный. После снижения на 10% заработная плата менеджера стала составлять 360 долл. Повышение на 10% этого заработка приведет к тому, что он станет получать 396 долл. ((360 × 10)/100) + 360), т.е. на 4 долл. меньше прежнего заработка. В первом случае снижение на 10% относилось к сумме 400 долл., а последующее повышение на 10% относилось к заработку в 360 долл.

Следовательно, относительные величины должны рассматриваться как ориентиры, дающие представление об изменениях фак-

тических данных или их соотношений, а не как заменители этих данных.

Одно из главных требований, предъявляемых при исчислении относительных величин, заключается в необходимости обеспечения сопоставимости сравниваемой величины и величины, принятой за базу сравнения. Прежде всего должна быть обеспечена сопоставимость по методологии расчета сравниваемых показателей, по степени охвата объектов исследуемой совокупности и другим существенным обстоятельствам (подробнее об этом см. гл. 8).

4.4. В чем состоит роль и значение средних величин?

В процессе обработки и обобщения первичных статистических данных по величине варьирующего признака возникает необходимость охарактеризовать исследуемую совокупность в целом, т.е. представить значения всей массы единиц неким одним характерным или типичным значением. Указанная задача решается с помощью средних величин, которые отражают основные свойства массы единиц изучаемой совокупности в отношении определенного признака.

Индивидуальные значения признака у отдельных единиц совокупности различны, и это различие связано с влиянием множества факторов, как общих для всех единиц совокупности, так и индивидуальных (случайных), присущих только отдельным единицам.

Например, уровень выработки рабочих за смену определяется организационно-техническим уровнем производства, продолжительностью рабочей смены, технологией производства. Однако в количественном отношении выработка каждого рабочего за смену различна, так как на нее кроме общих факторов влияют и индивидуальные особенности рабочего: уровень его квалификации, состояние трудовой дисциплины, образование и т.д.

Сущность средней состоит в том, что она отражает общие черты, закономерности, тенденции, присущие данной совокупности, погашая влияние индивидуальных (случайных) факторов, и поэтому является обобщающей характеристикой варьирующего признака единиц качественно однородной совокупности.

Средняя необходима для оценки индивидуальных значений признака у отдельных единиц совокупности, их грамотной интерпретации. Так, зная уровень рентабельности продукции на отдельном предприятии, инвестор может сделать обоснованный вывод о привлекательности вложений средств в данный вид деятельности

лишь при сравнении со средним уровнем рентабельности однотипных предприятий.

Средние величины играют важную роль в экономических расчетах, так как широко применяются в планировании, финансово-экономическом анализе деятельности предприятий, на их основе в сочетании с методом группировок выявляются взаимосвязи показателей и тенденции их развития. В связи с этим встает вопрос о надежности средней величины как обобщающей характеристики совокупности.

Средняя величина является надежной характеристикой свойств совокупности при условии, что индивидуальные величины, на основе которых она исчисляется, относятся к однородной совокупности, а само их число достаточно велико.

Однако одна и та же совокупность может по одному и тому же признаку быть однородной в одном отношении и неоднородной в другом, в зависимости от того, для какой цели применяется средняя. Например, автотранспортные предприятия общего пользования как отрасль являются однородной совокупностью и в этом смысле результаты работы отрасли (среднюю прибыль предприятий отрасли, среднюю рентабельность их активов и капитала, средние затраты на рабочую силу и т.д.) можно сравнивать с показателями других отраслей народного хозяйства, характеризуя экономическую конъюнктуру в экономике страны в целом.

Если же рассматривать каждое отдельное автотранспортное предприятие как хозяйствующий субъект, то их совокупность может быть крайне неоднородна, поскольку они обладают различными производственными мощностями, различны их организационно-технический уровень и организационно-правовые формы деятельности и т.д. Следовательно, результаты их работы могут варьировать в широких пределах. В подобных случаях в составе генеральной совокупности должны быть выделены качественно однородные группы и общая средняя должна быть дополнена групповыми средними, рассчитанными по этим однородным группам.

Между общей средней, рассчитанной для всей совокупности, и групповыми средними, рассчитанными для отдельных групп и подгрупп данной совокупности, имеются определенные различия. Эти различия обусловливаются наличием вариации в индивидуальных значениях признака у отдельных единиц совокупности. Групповые средние, дополняя и конкретизируя общую среднюю, дают возможность получить более разностороннюю характеристику изучаемой совокупности.

Приведем пример из статистики населения. Средний возраст населения России на начало 1999 г. составил 37,1 года, а доля лиц старше трудоспособного возраста в общей численности населе-

ния 20,8%. Однако если дополнить эти общие показатели по стране данными региональной статистики населения, то можно видеть, что в европейской части России проживает больше пожилых людей (для примера в табл. 4.3 приводятся данные по некоторым областям европейской части России).

Сравнение средних показателей России и отдельных областей, которые представляют собой группы, выделенные по региональному признаку, свидетельствует не только о более высоком показателе старения населения в областях европейской части России, но и о более высоком показателе демографической нагрузки[2].

Все виды средних величин, используемые в статистических исследованиях, подразделяются на две категории: 1) степенные средние; 2) структурные (позиционные) средние.

4.5. Как обосновать выбор вида степенной средней?

В практической статистике наибольшее распространение получили следующие виды степенных средних: средняя арифметическая, средняя гармоническая, средняя геометрическая и средняя квадратическая.

Назовем те общие правила, которыми следует руководствоваться при выборе вида степенной средней:

1. Формулируется цель, для которой рассчитывается средняя, и дается понятие определяющего показателя.
2. Приводится аналитическое выражение определяющего показателя.
3. Обосновывается формула расчета средней степенной, исходя из того, что величина определяющего показателя не изменится при замене индивидуальных значений признака у отдельных единиц совокупности средней величиной.

Отсюда можно сформулировать следующее определение: средней называется такая величина, которая не меняет значения определяющего показателя при подстановке ее в соответствующее аналитическое выражение вместо индивидуальных значений.

Допустим, что мы ставим задачу определения средней заработной платы некоторой совокупности работников. Очевидно, что ее величина может быть определена на основе фонда заработной платы:

[2] Показатель демографической нагрузки представляет собой число лиц старше трудоспособного возраста в расчете на 1000 человек трудоспособного возраста.

Таблица 4.3

Регионы	Показатели на начало 1999 г.	
	Средний возраст населения (в годах)	Доля лиц старше трудоспособного возраста в общей численности населения, (в %)
Россия	37,1	20,8
Области:		
Тульская	40,7	27,2
Рязанская	40,1	26,3
Тверская	39,7	25,6
Орловская	39,4	24,6
Ярославская	39,2	24,4

$$\text{Средняя заработная плата одного работника} = \frac{\text{Фонд заработной платы}}{\text{Число работников}}.$$

Фонд заработной платы является в данном случае определяющим показателем для средней заработной платы. Его величина получается путем суммирования заработков отдельных работников.

Обозначим через x_i заработную плату i-го работника, а n – число работников, тогда фонд заработной платы (Ф) будет равен:

$$\Phi = \sum_{i=1}^{n} x_i .$$

Эта запись представляет собой аналитическое выражение определяющего показателя. Чтобы найти формулу для расчета средней, нужно подставить в аналитическое выражение вместо индивидуальных значений их среднюю величину, т.е. $\bar{x}$, и тогда получим формулу простой средней арифметической:

$$\sum_{i=1}^{n} x_i = \bar{x} + \bar{x} + \bar{x} + \ldots + \bar{x} = n\bar{x},$$

$$\text{откуда } \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}.$$

Рассмотрим другой вариант имеющихся сведений, когда исходные данные для получения средней заработной платы представлены средними заработками отдельных групп работников ($\bar{x}_j$) и численностью работников соответствующей группы (n_j). В этом случае аналитическое выражение фонда заработной платы будет записано так:

$$\Phi = \sum_{j=1}^{k} \bar{x}_j \cdot n_j ,$$

где k — число различных групп (категорий) работников.

Воспользовавшись третьим правилом выбора средней, приравняем величины:

$$\sum_{j=1}^{k} \bar{x}_j \cdot n_j = \sum_{j=1}^{k} \bar{x}_0 \cdot n_j = \bar{x}_0 \cdot \sum_{j=1}^{k} n_j ,$$

где $\bar{x}_0$ — общая средняя для всей совокупности работников.

Из формулы аналитического выражения определяющего показателя находим формулу средней арифметической взвешенной:

$$\bar{x}_0 = \frac{\sum_{j=1}^{k} \bar{x}_j \cdot n_j}{\sum_{j=1}^{k} n_j} .$$

В данном примере «весом» служит численность работников соответствующей группы.

Использование средних взвешенных необходимо в тех случаях, когда отдельные групповые средние или отдельные варианты значений признака нельзя рассматривать как равноценные в формировании общей средней.

Если продолжить анализ формулы средней арифметической взвешенной, можно увидеть, что отношение $n_j / \sum_{j=1}^{k} n_j$ характеризует структуру изучаемой совокупности работников, иными словами, оценивает долю каждой группы работников в их общей численности. Обозначим эту долю через d_j, тогда формула средней арифметической взвешенной будет записана следующим образом:

$$\bar{x}_0 = \sum_{j=1}^{k} \bar{x}_j \cdot d_j .$$

Такая форма записи наглядно показывает, что на величину общей средней помимо групповых средних оказывает определенное влияние и структура совокупности.

Рассмотрим следующий пример. Портфель инвестора состоит из акций трех компаний. Их доходность равна соответственно 15, 18 и 20%, а доля в портфеле — 20, 45 и 35%. Какова средняя доходность портфеля?

В данном случае влияние доходности отдельных акций на среднюю доходность портфеля неравноценно, а потому расчет средней доходности портфеля следует выполнить по формуле средней арифметической взвешенной:

$$\bar{x}_{\text{портфеля}} = 15 \cdot 0{,}2 + 18 \cdot 0{,}45 + 20 \cdot 0{,}35 = 18{,}1\%.$$

Таким образом, по данным примера доходность портфеля составляет 18,1%.

Увеличивать доходность портфеля можно, либо приобретая акции более доходных компаний, либо увеличивая долю более доходных акций из уже имеющихся в портфеле. Например, изменим структуру портфеля так, чтобы доля акций первой компании составляла 10%, второй — 30% и третьей — 60%.

Тогда $\bar{x}_{\text{портфеля}} = 15 \cdot 0{,}1 + 18 \cdot 0{,}3 + 20 \cdot 0{,}6 = 18{,}9\%.$

4.6. Каковы основные свойства средней арифметической?

Средняя арифметическая величина обладает рядом свойств, имеющих значение при ее использовании.

Первое свойство состоит в том, что сумма отклонений индивидуальных значений признака от их средней величины равна нулю:

$$\sum_{i=1}^{n} (x_i - \bar{x}) = 0 .$$

Докажем это свойство:

$$\sum_{i=1}^{n} (x_i - \bar{x}) = \sum_{i=1}^{n} x_i - n\bar{x} = \sum_{i=1}^{n} x_i - n \frac{\sum_{i=1}^{n} x_i}{n} = 0 .$$

Второе свойство формулируется так: сумма квадратов отклонений индивидуальных значений признака от их средней величины есть величина минимальная, или:

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 < \sum_{i=1}^{n}(x_i - A)^2,$$

где $A = \bar{x} \pm \varepsilon$.

Рассмотрим отклонения x_i от величины, сколь угодно мало отличающейся от средней:

$$\sum_{i=1}^{n}(x_i - A)^2 = \sum_{i=1}^{n}[x_i - (\bar{x} \pm \varepsilon)]^2 = \sum_{i=1}^{n}[(x_i - \bar{x}) \pm \varepsilon]^2,$$

откуда очевидно, что:

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 < \sum_{i=1}^{n}[(x_i - \bar{x}) \pm \varepsilon]^2.$$

Третье свойство позволяет производить расчет средней не только с помощью абсолютных значений частоты повторения, но и с помощью относительных частот (показателей доли или удельного веса).

Это свойство формулируется следующим образом: если все частоты разделить на одно и то же число, средняя арифметическая останется без изменений.

4.7. Как обосновать выбор весов при расчете средней арифметической взвешенной?

При определении средней арифметической взвешенной немаловажное значение приобретает обоснование того показателя, который используется в качестве веса осредняемого признака. При выборе весов ориентируются прежде всего на сущность того показателя, для которого рассчитывается средняя. Поэтому в основе выбора веса лежит методика расчета того или иного экономического показателя.

Например, по данным об уровне рентабельности активов предприятий различных форм собственности следует определить средний по отраслям уровень рентабельности активов.

Для этого запишем принципиальную схему расчета этого показателя:

$$\text{Рентабельность активов} = \frac{\text{Прибыль за период}}{\text{Размер активов за период}} \cdot 100\%.$$

Таким образом, поскольку расчет рентабельности осуществляется на единицу активов, в качестве веса может использоваться только размер активов. Покажем расчет средней арифметической взвешенной на следующем примере.

Рентабельность активов акционерного общества (АО) характеризуется следующими данными:

Номер предприятия, входящего в состав АО	Величина активов (в тыс. руб.)	Рентабельность активов (в %)
1	5000	8,0
2	8000	9,5

Необходимо определить рентабельность активов по АО в целом.

В данном случае при расчете рентабельности активов АО вариантами являются рентабельности активов предприятия, входящих в состав АО, а весами — величины их активов:

$$\bar{R}_a = \frac{\sum R_{a_i} \cdot A_i}{\sum A_i} = \frac{\sum \Pi_i}{\sum A_i},$$

где $\bar{R}_a$ — средняя рентабельность активов АО;

R_{a_i} — рентабельность активов i-го предприятия;

A_i — величина активов i-го предприятия;

Π_i — прибыль i-го предприятия.

Как видно, выбор в качестве веса величины активов соответствует экономическому содержанию показателя рентабельности активов. Средний уровень рентабельности активов АО составит 8,9%:

$$\bar{R}_a = \frac{0{,}08 \cdot 5000 + 0{,}095 \cdot 8000}{5000 + 8000} \cdot 100 = 8{,}9\%.$$

Покажем принцип выбора весов еще на одном примере.

Финансирование инвестиций в основные средства за счет собственных средств по предприятиям акционерного общества за отчетный период характеризуется следующими данными:

Номер предприятия акционерного общества	Общая сумма инвестиций в основные средства (в млн. руб.)	Удельный вес собственных средств предприятий АО в общем объеме инвестиций (в %)
1	15,5	69,3
2	40,0	57,4
3	32,0	62,8

Необходимо определить средний удельный вес собственных средств в общем объеме инвестиций в основные средства по АО.

Исходя из экономического содержания показателя надо сравнить суммарные показатели собственных средств финансирования и общей суммы инвестиций в основные средства.

$$\text{Удельный вес собственных средств в финансировании основных средств} = \frac{\text{Объем собственных средств}}{\text{Общая сумма инвестиций в основные средства}} \cdot 100\%.$$

Тогда средний удельный вес собственных средств составит 61,5%:

$$\bar{x} = \frac{0{,}693 \cdot 15{,}5 + 0{,}574 \cdot 40 + 0{,}628 \cdot 32}{15{,}5 + 40 + 32} \cdot 100\% = 61{,}5\%.$$

В данном случае общая сумма инвестиций в основные средства выступает в качестве веса.

Чисто математически на этих примерах видно, что весом выступает знаменатель формулы расчета осредняемого показателя.

4.8. Что такое средняя гармоническая и средняя геометрическая?

Для понимания сущности средней гармонической обратимся к примеру. Предположим, что банк имеет данные о работе трех обменных пунктов валюты за день. Исходная информация может быть представлена в двух вариантах.

I вариант. По каждому обменному пункту известен курс доллара (рублей за один доллар) и объем продаж валюты (см. графу 2 и 3 табл. 4.4).

II вариант. По каждому обменному пункту известен курс доллара (рублей за один доллар) и выручка от продажи валюты (см. графу 4 табл. 4.4). Требуется рассчитать средний взвешенный курс доллара по трем обменным пунктам банка.

Прежде всего запишем формулу для определения курса доллара:

$$\text{Курс доллара (руб./долл.)} = \frac{\text{Выручка от продажи валюты}}{\text{Объем продаж}},$$

т.е. $x_i = B_i / q_i$.

Таблица 4.4

Номер обменного пункта	Валютный курс доллара (руб./долл.) x_i	Объем продаж (в тыс. долл.) q_i	Выручка от продажи валюты (в тыс. руб.) B_i
1	2	3	4
1	28,70	8,1	232,47
2	28,68	10,4	298,27
3	28,73	5,2	149,40
Итого		23,7	680,14

Если воспользоваться исходными данными I варианта, то в явном виде присутствует курс доллара и объем продаж валюты, а выручка от продажи валюты по каждому обменному пункту определяется так: $B_i = x_i \cdot q_i$.

Тогда расчет среднего курса доллара будет производиться по формуле средней арифметической взвешенной и он будет равен 28,70 руб./долл.:

$$\bar{x} = \frac{\sum_{i=1}^{n} B_i}{\sum_{i=1}^{n} q_i} = \frac{\sum x_i q_i}{\sum q_i} = \frac{28{,}70 \cdot 8{,}1 + 28{,}68 \cdot 10{,}4 + 28{,}73 \cdot 5{,}2}{23{,}7} =$$

$$= 28{,}70 \text{ руб./долл.}$$

Если же воспользоваться исходными данными II варианта, то из формулы курса валюты мы получим данные о выручке от продажи валюты, но в явном виде отсутствует объем продаж. Его величина по каждому пункту определяется отношением:

$$q_i = B_i / x_i.$$

И тогда формула для расчета среднего курса будет записана так:

$$\bar{x}_{\text{гарм}} = \frac{\sum_{i=1}^{n} B_i}{\sum_{i=1}^{n} \frac{B_i}{x_i}},$$

где $\bar{x}_{\text{гарм}}$ — средняя гармоническая;

n — количество вариантов значений признака.

В этом случае мы пришли к формуле средней гармонической взвешенной.

Результат расчета по формуле средней гармонической взвешенной будет тождествен расчету по средней арифметической взвешенной:

$$\bar{x} = \frac{232{,}47 + 298{,}27 + 149{,}40}{\frac{232{,}47}{28{,}70} + \frac{298{,}27}{28{,}68} + \frac{149{,}40}{28{,}73}} = 28{,}70 \text{ руб./долл.}$$

Таким образом, выбор вида средней в данном случае обусловлен составом имеющихся исходных данных.

Формула простой средней гармонической (например, при равенстве выручки от продаж валюты в обменных пунктах) записывается следующим образом:

$$\bar{x}_{\text{гарм}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}.$$

Средняя геометрическая применяется для исчисления среднего коэффициента роста в рядах динамики.

Например, имеются следующие данные о выпуске продукции по кварталам отчетного года.

Показатель	Кварталы года			
	I	II	III	IV
Выпуск продукции (в тыс. руб.)	100	120	125	122
Условные обозначения	y_0	y_1	y_2	y_3
Коэффициент роста (снижения) выпуска продукции по сравнению с предыдущим кварталом	—	1,2	1,04	0,976
Условные обозначения		$K_1 = y_1/y_0$	$K_2 = y_2/y_1$	$K_3 = y_3/y_2$

Необходимо определить средний квартальный коэффициент роста выпуска продукции.

Определяющим показателем по отношению к этим цепным коэффициентам роста будет базисный темп роста, который показывает изменение показателя выпуска продукции за весь период, т.е. $K_0 = y_3/y_0$. Тогда, используя взаимосвязь цепных и базисных коэффициентов роста, охватывающих один и тот же период времени, и правила выбора формулы средней величины, можно записать:

$$K_{3/0} = \frac{y_3}{y_0} = k_1 \cdot k_2 \cdot k_3 = \bar{k} \cdot \bar{k} \cdot \bar{k} = \bar{k}^3,$$

откуда $\bar{k} = \sqrt[3]{k_1 \cdot k_2 \cdot k_3}$;

$$\bar{k} = \sqrt[3]{1{,}2 \cdot 2{,}04 \cdot 0{,}976} = \sqrt[3]{\frac{122}{100}} = 1{,}068.$$

Следовательно, в общем виде формула средней геометрической простой будет иметь вид:

$$\bar{K} = \sqrt[n-1]{K_1 \cdot K_2 \cdot \ldots \cdot K_{n-1}},$$

где n — число уровней временного ряда;

$K_1 \cdot K_2 \cdot \ldots \cdot K_{n-1}$ — цепные коэффициенты роста.

Формула средней геометрической простой применяется для расчета среднего коэффициента роста, если промежутки времени, к которым относятся коэффициенты роста, одинаковы.

Если известные средние коэффициенты роста относятся к разным по продолжительности временным отрезкам m_1, m_2,..., m_k, то общий средний коэффициент роста ($K_{\text{общ}}$) за весь рассматриваемый период определяется по формуле средней геометрической взвешенной, которая имеет вид:

$$\bar{K}_{\text{общ}} = \sqrt[\sum_{j=1}^{l} m_j]{\bar{K}_1^{m_1} \cdot \bar{K}_2^{m_2} \cdot \ldots \bar{K}_l^{m_l}},$$

где m_j — продолжительность периода, к которому относится средний коэффициент роста;

$\bar{K}_1^{m_1} \cdot \bar{K}_2^{m_2} \cdot \ldots \bar{K}_l^{m_l}$ — средние коэффициенты роста в отдельные, разные по продолжительности периоды.

Например, известны среднегодовые темпы роста душевого дохода в США:

Годы	1913—1950 гг.	1950—1973 гг.	1973—1992 гг.
Среднегодовой темп роста (в %)	101,6	102,4	101,4

Требуется определить среднегодовой темп роста душевого дохода в США в рассматриваемом периоде. Его величина составит 101,8%:

$$\overline{K}_p = \sqrt[37+23+19]{1{,}016^{37} \cdot 1{,}024^{23} \cdot 1{,}014^{19}} = 101{,}8\%.$$

Общая формула степенной средней записывается так:

$$\overline{x} = \sqrt[k]{\frac{\sum_{i=1}^{n} x_i^k}{n}},$$

где k — степень средней..

При $k = 1$ получаем формулу средней арифметической; $k = 0$ соответствует средняя геометрическая; при $k = -1$ получаем среднюю гармоническую, при $k = 2$ расчет ведется по формуле средней квадратической.

Если разные виды средних величин рассчитать по одним исходным данным, то между численными значениями этих средних будет иметь место следующее соотношение:

$$\overline{x}_{\text{гарм}} < \overline{x}_{\text{геом}} < \overline{x}_{\text{арифм}} < \overline{x}_{\text{квадр}}.$$

Это соотношение называется правилом мажорантности средних величин.

4.9. Что такое структурные средние и как их определять по несгруппированным данным?

Для расчета степенных средних, как уже было показано выше, в большинстве случаев необходимо располагать данными не только о вариантах значений признака в исследуемой совокупности, но и о значимости, или весомости отдельных вариантов значений.

Поэтому при условии недостаточности исходных данных, которая в ряде случаев объективно может возникать при сборе информации (например, ограничения в связи с так называемой «коммерческой тайной»), предпочтение отдается структурным, или позиционным, средним — медиане и моде.

Медианой называется значение варьирующего признака, которое делит ранжированный ряд данных на две равные части: 50% единиц исследуемой совокупности будут иметь значения признака меньше, чем медиана, а 50% — значения признака больше, чем медиана.

При определении медианы по несгруппированным (первичным) данным сначала необходимо расположить их в возрастающем порядке (ранжировать). Затем нужно определить «позицию» медианы или определить номер той единицы, значение признака у которой будет соответствовать медиане:

$$N_{Me} = (n + 1)/2,$$

где n — число единиц в исследуемой совокупности.

Например, по предприятиям малого бизнеса, занимающимся торгово-посреднической деятельностью, известны данные о коэффициенте покрытия[3].

Номер предприятия	1	2	3	4	5	6	7
Коэффициент покрытия	1,12	1,22	1,00	1,16	1,08	1,22	1,00

8	9	10	11	12	13	14	15
1,44	1,00	1,14	1,00	1,35	1,47	1,00	1,25

Расположим коэффициенты покрытия в возрастающем порядке: 1,00; 1,00; 1,00; 1,00; 1,00; 1,08; 1,12; 1,14; 1,16; 1,22; 1,22; 1,25; 1,35; 1,44; 1,47.

Поскольку мы располагаем данными по пятнадцати предприятиям, медианой будет восьмое значение в ранжированном ряду:

$$N_{Me} = (15 + 1)/2 = 8.$$

Таким образом, значению медианы будет соответствовать коэффициент покрытия 1,14.

Если бы мы располагали данными по четному числу предприятий, то номер медианы был бы дробным. Например, при n=16 номер медианы был бы равен 8,5 [(16 + 1)/2]. А это означало бы, что медиана представляет собой среднюю арифметическую из восьмого и девятого значений признака в ранжированном ряду.

Модой называют то значение признака, которое наиболее часто встречается в данной совокупности. Так, по данным приводимого примера наиболее часто для данной группы предприятий малого бизнеса повторяется коэффициент покрытия, равный 1,00.

Могут быть случаи, когда два и даже более значений признака повторяются одинаковое число раз. В случае наличия двух модальных значений говорят о бимодальном распределении признака, при наличии нескольких модальных значений — о мультимодальном распределении. Наличие двух и более модальных значений может означать неоднородность исследуемой совокупности.

При отсутствии повторяющихся значений признака в совокупности первичных данных для определения моды необходимо предварительно выполнить группировку, в результате которой мы будем иметь дело с интервальным рядом распределения. Расчет моды и медианы по сгруппированным данным показан в гл. 5.

[3] Коэффициент покрытия = $\frac{\text{Оборотные средства}}{\text{Краткосрочные обязательства}}$.

Глава 5
СТАТИСТИЧЕСКИЕ РАСПРЕДЕЛЕНИЯ И ИХ ОСНОВНЫЕ ХАРАКТЕРИСТИКИ

5.1. Что такое вариация признака в совокупности и в чем значение ее изучения?

Составной частью сводной обработки данных статистического наблюдения является построение рядов распределения. Его цель состоит в выявлении основных свойств и закономерностей исследуемой статистической совокупности. В зависимости от того, является ли признак, взятый за основу группировки, качественным или количественным, различают соответственно два типа рядов распределения — атрибутивные и вариационные. Ряды распределения, построенные по качественным признакам, называют ***атрибутивными***. Примером атрибутивных рядов может служить распределение населения по полу, характеру труда, национальности, профессии и т.д. Ряды распределения, построенные по количественному признаку, называют ***вариационными***. Величины того или иного количественного признака у отдельных единиц совокупности более или менее различаются между собой. Такое различие в величине признака носит название вариации.

Числовые значения признака, встречающиеся в данной совокупности, называют вариантами значений. Наличие различий в величине признака у отдельных единиц совокупности обусловлено влиянием большого числа факторов, формирующих уровень признака. Часто эти факторы могут оказывать разнонаправленное воздействие на уровень анализируемого показателя. Например, снижение цен на строительные материалы может привести к уменьшению цен на 1 м² жилья, а увеличение спроса на жилье может повысить эту цену и т.д. В результате совместного влияния различных факторов и складывается цена 1 м² муниципального жилья в определенное время. Но есть, например, и такой фактор, как экологическая обстановка в разных районах, которая также обусловливает вариацию цен 1 м² жилья в разных районах города. Поэтому при изучении вариации различных показателей можно выделить две группы факторов, формирующих уровень признака в исследуемой совокупности единиц и обусловливающих существование различий в величине признака у отдельных единиц. *Первую группу составляют факторы, общие для всех единиц*

изучаемой совокупности. Во вторую группу входят факторы, свойственные конкретным единицам совокупности и определяющие их индивидуальные особенности.

Изучение характера и степени вариации признаков у отдельных единиц, составляющих изучаемую совокупность, является важнейшим вопросом всякого статистического исследования. Управление процессом развития в желаемом направлении требует определения роли не только каждой из выше названных групп факторов в вариации тех или иных признаков, но и роли отдельных факторов соответствующих групп. Для решения такой задачи в статистике применяются специальные методы исследования вариации, основанные на использовании системы показателей, с помощью которой измеряется вариация.

Собранные в процессе наблюдения данные без какой-либо их систематизации образуют так называемый ***первичный ряд*** данных. При наличии достаточно большого количества единиц в обследованной совокупности первичный ряд становится трудно обозримым и непосредственное рассмотрение его не дает представления о распределении единиц по величине признака в совокупности. Первым шагом в упорядочении первичного ряда является его ***ранжирование,*** т.е. расположение всех вариантов ряда в возрастающем (или убывающем) порядке.

Ранжирование данных позволяет сразу увидеть наименьшее и наибольшее значения признака в совокупности, определить расстояние между крайними значениями признака, а также выделить наиболее часто повторяющиеся значения в обследуемой совокупности. Использование ранжированного ряда также позволяет легко разделить все данные по группам.

Число повторений отдельных вариантов значений признаков называют частотой повторения. В дальнейшем частоту повторения значения признака будем обозначать как f_i, а сумму частот, равную объему изучаемой совокупности, как $\sum_{i=1}^{k} f_i$, или $n(\sum_{i=1}^{k} f_i = n)$, где k — число вариантов значений признака.

По характеру вариации различают дискретные и непрерывные признаки. ***Дискретные признаки*** отличаются друг от друга на некоторую конечную величину, т.е. даны в виде прерывных чисел, например: тарифный разряд рабочих, число детей в семье, число рабочих на предприятии и т.д. ***Непрерывные признаки*** могут отличаться один от другого на сколь угодно малую величину и в определенных границах принимать любые значения, например: заработная плата рабочих, размер среднедушевого денежного дохода, стоимость основных фондов предприятия и т.д.

5.2. В чем состоят различия в построении рядов распределения с дискретным и непрерывным характером вариации признака?

Для построения дискретного ряда с небольшим числом вариантов достаточно перечислить все встречающиеся варианты значений признака, обозначаемые через x_i, а затем подсчитать частоту повторения каждого варианта f_i (например, распределение рабочих по разрядам, студентов по успеваемости и т.п.). Ряд распределения принято оформлять в виде таблицы. Например, распределение рабочих участка по квалификации характеризуется данными табл. 5.1.

Таким образом, ряд первичных данных, характеризующих квалификацию 20 рабочих, заменен коротким рядом, состоящим из пяти групп. Вместо абсолютного числа рабочих, имеющих определенный разряд, можно установить долю рабочих этого разряда. Частоты, представленные в относительном выражении, называют ***частостями*** и обозначаются w_i (графа 3 табл. 5.1):

$$W_i = \frac{f_i}{\sum_{i=1}^{k} f_i}.$$

Частости могут быть выражены в долях единицы или в процентах. Замена частот частостями позволяет сопоставлять вариационные ряды с различным числом наблюдений.

В тех случаях, когда число вариантов дискретного признака достаточно велико, а также при анализе вариации непрерывного признака, когда значения признака у отдельных единиц могут вообще не повторяться, строятся ***интервальные ряды распределения***. Интервал указывает определенные пределы значений варьирующего признака и обозначается нижней и верхней границами интервала, т.е. нижняя граница интервала показывает то значение, с которого начинается данная группа, а верхняя – значение, на котором эта группа заканчивается. Такие распределения наиболее распространены в практике статистической работы.

Таблица 5.1

Тарифный разряд рабочего (X_i)	Число рабочих, имеющих соответствующий разряд (f_i)	Частость (w_i)	Накопленная частота (S_i)
1	2	3	4
2	1	0,05	1
3	5	0,025	6
4	8	0,40	14
5	4	0,20	18
6	2	0,10	20
Итого	20	1,00	

При построении интервальных рядов распределения необходимо прежде всего установить число групп (интервалов), на которые следует разбить все единицы изучаемой совокупности. При группировке внутри однородных совокупностей появляется возможность применения равных интервалов, величина которых зависит от вариации признака в совокупности и от количества обследованных единиц.

Определение величины интервала h для построения вариационного ряда с равными интервалами производится следующим образом:

1) вычисляется разность между максимальным и минимальным значениями признака первичного ряда (определяется размах вариации, R):

$$R = x_{max} - x_{min};$$

2) размах вариации делится на число групп k, т.е. $h = R/k$.

Число групп приближенно определяется по формуле Стэрджесса: $k \approx 1+3{,}322\lg n$, где n — общее число изучаемых единиц совокупности.

Указанное выражение почти всегда оказывается дробной величиной, которую округляют до целого числа, поскольку количество групп не может быть дробным (как правило, лучше округлять в меньшую сторону).

Величина интервала должна определяться в соответствии с точностью данных наблюдения: если исходные данные представлены целыми числами, то рассчитанная величина интервала округляется до ближайшего целого числа; если данные представлены с точностью до 0,1, то величина интервала округляется до целых с десятыми и т.д. (здесь округление производится в большую сторону).

Знание величины интервала позволяет определить границы всех интервалов ряда распределения. Нижнюю границу первого интервала целесообразно принимать равной минимальному значению признака. Имеются и предложения иного рода. Так, нижнюю границу первого интервала рекомендуют определять путем вычитания из минимального значения признака половины величины интервала.

В каждый интервал включаются варианты, числовые значения которых больше или равны нижней границе и меньше или равны верхней границе, и затем подсчитывается число единиц в каждом интервале. При построении интервальных рядов для непрерывных признаков имеет место совпадение верхних границ предшествующих интервалов и нижних границ следующих за ними интервалов. Здесь должны даваться пояснения, в какой интервал относить единицы совокупности, числовые значения признака у которых совпадают с одной из этих границ.

В интервальных рядах распределения дискретных признаков отнесение единиц совокупности в ту или иную группу не вызывает затруднений, так как между верхней границей одного интервала и нижней границей смежного интервала существует разрыв. Иная стадия возникает в рядах распределения с непрерывным признаком.

Рассмотрим построение ряда распределения по первичным данным о размере прибыли 20 коммерческих банков региона за год (в млн руб.):

x_i – 3,7; 4,3; 6,7; 5,6; 5,1; 8,1; 4,6; 5,7; 6,4; 5,9; 5,2; 6,2; 6,3; 7,2; 7,9; 5,8; 4,9; 7,6; 7,0; 6,9.

Поскольку варианты значений признака не повторяются, строим интервальный ряд распределения.

Определяем количество групп интервального вариационного ряда: $k = 1 + 3{,}322\lg 20 = 1 + 3{,}322 \times 1{,}301 = 5{,}32$. Округляя, получим число групп, равное 5. Тогда величина интервала составит 0,9 млн руб. ($h = (8{,}1 - 3{,}7)/5 = 0{,}9$). В результате подсчета числа банков в каждой группе получим ряд распределения банков по величине полученной прибыли за год (см. графы 1, 2 табл. 5.2). Значения признака у отдельных единиц совпали с границами интервалов (например, значения 4,6 и 6,4 млн руб.). Так как минимальное значение признака (3,7 млн руб.) совпадает с нижней границей первого интервала и включается в этот интервал, значения 4,6 и 6,4 млн руб. следует включать в тот интервал, нижняя граница которого совпадает с указанным значением (например, значение признака 4,6 млн руб. включается во второй интервал, а значение признака 6,4 млн руб. — в четвертый).

Для пояснения в первой строке графы 1 табл. 5.2 поставлен знак «минус». Этот знак соответствует принципу «исключая» и означает, что величины признака, совпадающие с верхней границей интервала, в этот интервал не включаются, а попадают в следующий интервал. Если ставится знак «плюс», это соответствует принципу «включая» и означает, что величины признака, совпадающие с верхней границей интервала, включаются в этот интервал.

Таблица 5.2

Размер прибыли	Число банков (частота) f_i	Накопленная частота S_i	X_i'	$X_i' \times f_i$	Удельный вес группы (частость)		Накопленные частости	
					По числу банков W_i	По объему прибыли d_i	По числу банков p_i	По объему прибыли q_i
1	2	3	4	5	6	7	8	9
3,7–4,6	2	2	4,15	8,30	0,1	0,0682	0,1	0,0682
4,6–5,5	4	6	5,05	20,20	0,2	0,1660	0,3	0,2342
5,5–6,4	6	12	5,95	35,70	0,3	0,2933	0,6	0,5275
6,4–7,3	5	17	6,85	34,25	0,25	0,2815	0,85	0,8090
7,3–8,2	3	20	7,75	23,25	0,15	0,1910	1,00	1,0000
Итого	20			121,70	1,00	1,0000		

* $W_i = \frac{f_i}{\sum f_i}$; $W_1 = \frac{2}{20} = 0{,}1$; $W_2 = \frac{4}{20} = 0{,}2$ и т.д.

** $d_i = \frac{X_i f_i}{\sum X_i f_i}$; $d_1 = \frac{8{,}30}{121{,}70} = 0{,}0682$; $d_2 = \frac{20{,}20}{121{,}70} = 0{,}1660$ и т.д.

5.3. Какие группы обобщающих показателей используют для характеристики особенностей рядов распределения?

Для целей анализа и сравнительной характеристики различных рядов распределения применяются обобщающие показатели вариационного ряда. Система таких показателей может быть наглядно представлена при сравнении особенностей нескольких рядов распределения.

Так, на рис. 5.1 кривые распределения 1 и 2 имеют одинаковый размах вариации и характер распределения частот, но отличаются величиной варьирующего признака, являющегося центром группирования (отмечен на оси x).

Характеристики центра группирования, таким образом, составляют одну из групп обобщающих показателей. К ним относятся средняя арифметическая, мода и медиана.

Представленные на рис. 5.2 кривые распределения 3 и 4 имеют один и тот же центр группирования и симметричное расположение частот вокруг него, но отличаются пределами вариации. Следова-

тельно, кроме показателей центра группирования, для характеристики особенностей распределения необходимо иметь представление о степени колеблемости признака и знать показатели степени вариации. Эти две группы показателей — показатели центра распределения и вариации признака — имеют особое значение при принятии решений в управлении.

Ряды распределения могут иметь один и тот же центр группирования, одинаковые пределы варьирования признака, симметричный характер расположения частот, но разную степень вытянутости вдоль оси ординат, которая характеризуется показателями эксцесса. И наконец, сравнение различных распределений показывает, что они могут отличаться характером распределения частот относительно центра; степень отклонения распределения частот от симметричной формы характеризуется показателями асимметрии. Показатели эксцесса и асимметрии характеризуют форму распределения.



Рис. 5.1. Кривые распределения с разными центрами группирования

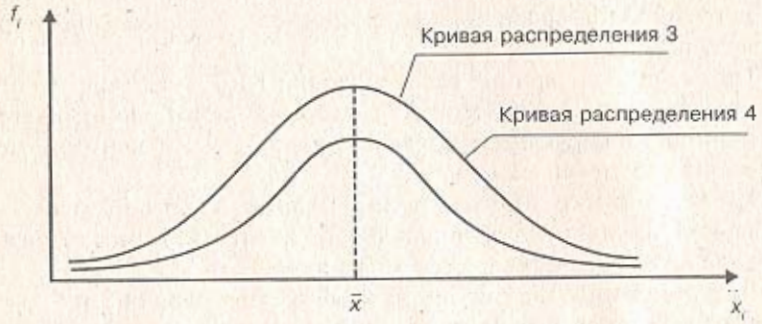


Рис. 5.2. Кривые распределения с различной вариацией признака

Таким образом, в зависимости от характеризуемых особенностей распределения обобщающие показатели можно разбить на три группы:

- показатели центра распределения (центра группирования) позволяют сделать вывод о центральном или наиболее общем значении, найденном для совокупности данных;
- показатели степени вариации показывают, как данные распределены вокруг средней;
- показатели формы распределения иллюстрируют степень скошенности и уровень островершинности или плосковершинности распределения.

5.4. Как графически представить ряд распределения?

Графическое изображение рядов распределения облегчает их анализ и позволяет судить о форме распределения. Начнем с графического изображения дискретного ряда с помощью ***полигона распределения***. Для его построения на оси абсцисс отмечают точки, соответствующие величине вариантов значений признака, из них восстанавливаются перпендикуляры, длина которых соответствует частоте (частости) этих вариантов по принятому масштабу на оси ординат. Вершины перпендикуляров в последовательном порядке соединяются отрезками прямых. Для замыкания полигона крайние вершины соединяются с точками на оси абсцисс, отстоящими на одно деление в принятом масштабе от x_{max} и x_{min}. Такое построение полигона облегчает зрительное восприятие его графического изображения.

Воспользуемся данными, приведенными в табл. 5.2, для построения полигона (рис. 5.3).

Для графического изображения интервальных вариационных рядов применяется гистограмма. Она строится так: на оси абсцисс откладываются равные отрезки, которые в принятом масштабе соответствуют величине интервалов вариационного ряда. На отрезках строят прямоугольники, площади которых пропорциональны частотам (или частостям) интервала.

На рис. 5.4 изображена гистограмма ряда распределения банков по размеру прибыли (по данным табл. 5.2). Гистограмма может быть

преобразована в полигон распределения, если середины верхних сторон прямоугольников соединить отрезками прямых. Две крайние точки прямоугольников замыкаются по оси абсцисс на середины интервалов, в которых частоты (частости) равны нулю. При построении гистограммы для вариационного ряда с неравными интервалами следует по оси ординат наносить показатели плотности интервалов (абсолютные или относительные). В этом случае высоты прямоугольников гистограммы будут соответствовать величине плотности распределения.

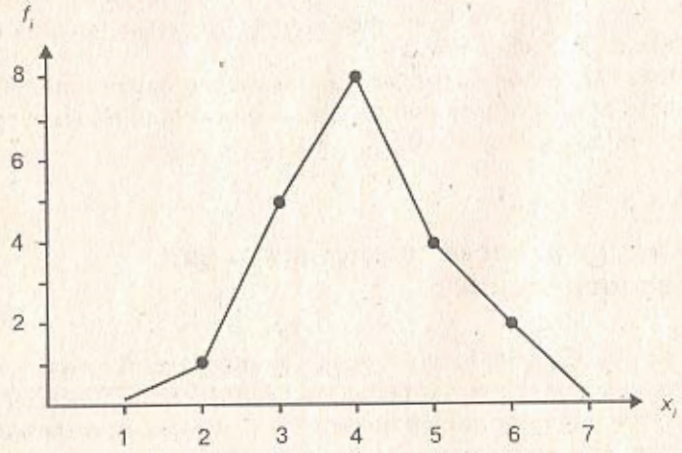


Рис. 5.3. Распределение рабочих участка по квалификации: x_i — тарифный разряд; f_i — число рабочих

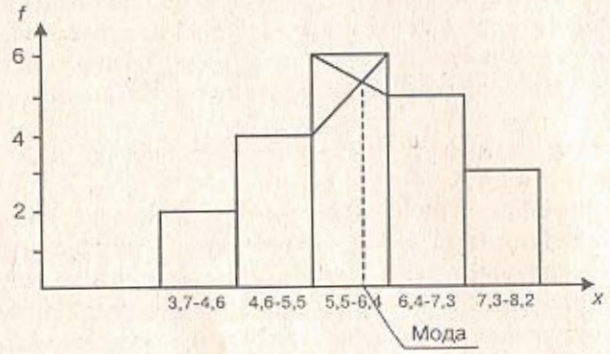


Рис. 5.4. Распределение коммерческих банков по размеру прибыли (x — размер прибыли (в млн руб.), f — число банков)

Для определения плотности интервала вычисляют отношение частот, или частостей, к величине интервала:

абсолютная плотность — $m_i^a = f_i / h_i$;

относительная плотность — $m_i^o = W_i / h_i$.

Эти показатели используют также для преобразования интервалов, что бывает необходимо при сравнительной оценке данных, собранных по различным совокупностям и по-разному обработанных.

При увеличении числа наблюдений из одной и той же совокупности увеличивается число групп интервального ряда, что соответственно приводит к уменьшению величины интервала. При этом ломаная линия имеет тенденцию превращения в плавную кривую, которую называют ***кривой распределения***. Эта кривая характеризует в обобщенном виде вариацию признака и закономерности распределения частот внутри совокупности.

В ряде случаев для изображения вариационных рядов используется ***кумулятивная кривая*** (кумулята). Для ее построения надо рассчитать накопленные частоты или частости. Накопленные частоты показывают, сколько единиц совокупности имеют значения признака не больше, чем рассматриваемое значение, и определяются последовательным суммированием частот интервалов. Построим кумулятивную кривую по данным табл. 5.2 о распределении банков по размеру прибыли. (Накопленные частоты рассчитаны в графе 3 табл. 5.2.) При построении кумуляты интервального ряда распределения нижней границе первого интервала соответствует частота, равная нулю, а верхней границе — вся частота данного интервала. Верхней границе второго интервала соответствует накопленная частота, равная сумме частот первых двух интервалов, и т.д. (рис. 5.5).

Изображение вариационного ряда в виде кумуляты особенно удобно при сравнении вариационных рядов, а также в экономических исследованиях, в частности, для анализа концентрации производства.

5.5. В чем состоят особенности расчета средней арифметической в вариационных рядах?

Для характеристики среднего значения признака в вариационном ряду используются средняя арифметическая, мода и медиана. Общие понятия о средних величинах и их свойствах даны в гл. 4. В данном параграфе рассмотрим расчет показателей центра распределения для вариационных рядов.

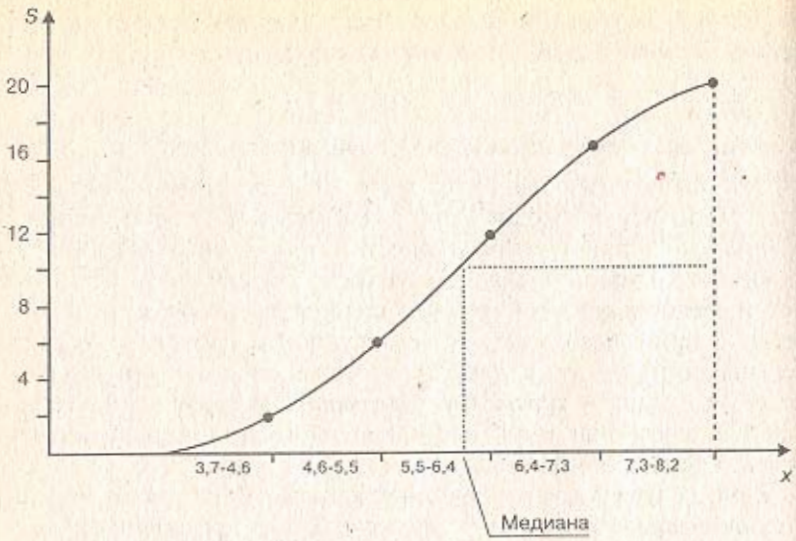


Рис. 5.5. Кумулята ряда распределения банков по размеру прибыли (x — размер прибыли (в млрд руб.), S — накопленные частоты)

Средняя арифметическая для дискретного ряда распределения рассчитывается по формуле:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i},$$

где x_i — варианты значений признака;

f_i — частота повторения данного варианта.

В интервальном ***вариационном ряду*** расчет ***средней арифметической*** основан на использовании центральных значений интервалов:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i^1 f_i}{\sum_{i=1}^{k} f_i};$$

$$x_i^1 = \frac{x_i^{\text{н}} + x_i^{\text{в}}}{2},$$

где x_i^1 — центральное значение i-го интервала;

$x_i^{\text{н}}$ и $x_i^{\text{в}}$ — соответственно нижняя и верхняя границы i-го интервала.

Пример расчета средней арифметической для интервального вариационного ряда приведен далее, в табл. 5.2, ее величина составляет 6,085 млн руб. ($\bar{x}$ = 121,70/20 = 6,085).

Произведение $x_i f_i$ или $x_i^1 f_i$ характеризует объем изучаемого показателя в i-й группе. В графе 5 табл. 5.2 представлен объем прибыли, полученной всеми банками соответствующей группы. Так, два банка первой группы получили 8,30 млн руб. прибыли за год, четыре банка второй группы – 20,20 млн руб. прибыли и т.д. Сумма произведений $\sum_{i=1}^{k} x_i^1 f_i$ характеризует общий объем исследуемого показателя в совокупности. В заданном примере суммарный объем прибыли 20 банков за год составил 121,70 млн руб.

В графах 6 и 7 табл. 5.2 приведены показатели удельного веса (доли) каждой группы по числу единиц (W_i) и по объему прибыли (d_i).

Сопоставление объема изучаемого показателя группы (в долях или процентах к итогу) с числом единиц этой группы (в долях или процентах к итогу) позволяет характеризовать роль отдельных групп в формировании объема изучаемого показателя. Такое сопоставление дает возможность изучать процесс концентрации.

5.6. Как оценить степень неравномерности распределения признака?

Для оценки степени неравномерности распределения объема изучаемого признака между группами абсолютные показатели числа единиц в группе и размера изучаемого признака (в данном примере это прибыль банка) следует выразить в относительных показателях (в долях или процентах к итогу). Затем рассчитывают два ряда накопленных частостей (см. графы 8 и 9 табл. 5.2).

Используя данные этой таблицы, представим графически соотношения накопленных частостей (рис. 5.6). На горизонтальной оси (оси абсцисс) расположим накопленные частости, рассчитанные по числу банков (p_i), а на оси ординат наносим накопленные частости, определенные по объему прибыли (используем данные двух последних граф табл. 5.2).

Кривая, отмеченная на графике пунктирной линией, изображает линию фактического распределения и носит название кривой Лоренца. Она характеризует степень неравномерности распределения общего объема показателя между группами, в данном примере объема прибыли между группами банками. Сплошная линия (ли-

ния равномерного распределения) означает, что определенному проценту банков соответствует такой же процент по объему прибыли, т.е. 10% банков должно было бы иметь 10% прибыли, 20% банков — 20% прибыли и т.д. Это могло бы быть при условии, что общая сумма прибыли разделена поровну между банками, а график представлял бы собой прямую, проходящую через начало координат под углом 45° на оси абсцисс. Тогда по отклонению кривой Лоренца от линии равномерного распределения можно сулить о степени неравномерности фактического распределения прибыли между банками.

Чем более удалена линия фактического распределения от линии равномерного рапределения, тем более неравномерно распределен объем изучаемого показателя между единицами (группами) статистической совокупности. Графический метод может быть использован для сопоставления кривых Лоренца, характеризующих концентрацию данного показателя в разных совокупностях, что позволяет сравнить степень концентрации в этих совокупностях, а также сопоставить кривые Лоренца данного показателя за разные периоды, что позволяет выявить тенденции в неравномерности распределения объема признака между группами. Такие сопоставления достаточно распространены в социально-экономической статистике,

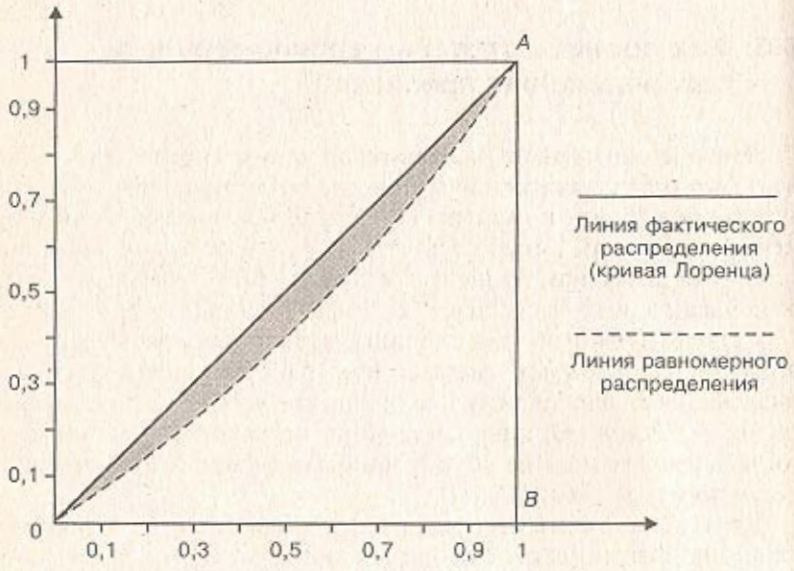


Рис. 5.6. Графическая интерпретация степени концентрации прибыли банков

например, изучение распределения объема денежных доходов между различными группами населения, анализ степени концентрации банковского капитала, сравнение концентрации объема производства в различных отраслях промышленности и т.д.

Площадь между линией равномерного распределения и кривой Лоренца, заштрихованная на рис 5.6, может послужить для расчета одного из показателей концентрации, называемого индексом Джини (обозначим его $K_Д$):

$$K_д = S_ч / S_{OAB},$$

где $S_ч$ — площадь между линией равномерного и фактического распределения, (заштрихована на графике);

S_{OAB} — площадь треугольника OAB, которая при расчете накопленных частостей в долях единицы равна 0,5.

Величина $K_Д$ изменяется в пределах от 0 до 1. Для равномерного распределения индекс Джини равен нулю, в условиях полной концентрации он равен 1. Чем больше степень концентрации, тем ближе величина индекса Джини приближается к 1. В Российском статистическом ежегоднике в качестве одного из социально-экономических индикаторов уровня жизни населения приводится индекс Джини, устанавливающий степень отклонения фактического распределения объема доходов населения от линии равномерного распределения. Чтобы выполнить анализ этого показателя для справок, приводятся коэффициенты концентрации доходов в отдельных странах.

5.7. Как определить моду и медиану в интервальном ряду распределения?

В отличие от средней арифметической, рассчитываемой на основе использования всех вариантов значений признака, мода и медиана характеризуют величину варианта, занимающего определенное положение в ранжированном вариационном ряду. ***Медиана (Me)*** соответствует варианту, стоящему в середине ранжированного ряда. Положение медианы определяется ее номером $N_{Me} = (n + 1)/2$, где n — число единиц в совокупности.

Используем данные примера, приведенные в табл. 5.1 (см. параграф 5.2), для определения медианы и моды. Позиция значения признака, соответствующего медиане, определяется ее номером. В данном примере $N_{Me} = (20 + 1)/2 = 10{,}5$, т.е. медиана равна средней арифметической из 10-го и 11-го значений признака. По накопленным частотам определяем, что 10-я и 11-я единицы ряда

имеют величину признака, равную четвертому разряду, т.е. в примере медиана равна четвертому разряду.

Мода (Mo) — наиболее часто встречающееся значение признака в совокупности — для данного ряда распределения также равна четвертому разряду, так как этому разряду соответствует максимальная частота, равная 8.

В интервальном ряду распределения сразу можно указать только интервал, в котором будут находиться мода или медиана. Для определения их величины используются следующие формулы[4]:

$$Me = x_{Me} + h\frac{\frac{n+1}{2} - S_{(-1)}}{f_{Me}},$$

где x_{Me} — нижняя граница медианного интервала;
h — величина интервала;
$S_{(-1)}$ — накопленная частота интервала, предшествующего медианному;
f_{Me} — частота медианного интервала.

Используя данные примера, приведенные в табл. 5.2, рассчитаем медиану. По накопленным частотам (графа 3) определяем, что медиана находится в интервале 5,5—6,4. Тогда

$$Me = 5{,}5 + 0{,}9\frac{\frac{20+1}{2} - 6}{6} = 6{,}175 \text{ млн руб.}$$

Таким образом, 50% банков имеют прибыль менее 6,175 млн руб., а 50% банков — более 6,175 млн руб.

Наибольшая частота соответствует также интервалу 5,5—6,4, т.е. мода должна находиться в этом интервале. Ее величину определяем по формуле:

$$Mo = x_{Mo} + \frac{f_{Mo} - f_{(-1)}}{[f_{Mo} - f_{(+1)}] + [f_{Mo} - f_{(-1)}]},$$

где x_{Mo} — начало модального интервала;
f_{Mo} — частота, соответствующая модальному интервалу;
$f_{(-1)}$ — предмодальная частота;
$f_{(+1)}$ — послемодальная частота.

[4] Формула медианы выведена, исходя из предположения о том, что плотность внутри интервала остается постоянной, т.е. данные равномерно распределены внутри интервала.

Приведенная формула моды может быть использована в вариационных рядах с равными интервалами:

$$Mo = 5{,}5 + \frac{6-4}{[6-4]+[6-5]} = 6{,}10 \text{ млн руб.}$$

Таким образом, в данной совокупности банков наиболее часто встречается размер прибыли 6,10 млн руб.

Медиану и моду можно также определить графически. Медиана определяется по кумуляте (см. рис. 5.5). Для ее определения высоту наибольшей ординаты, которая соответствует общей численности совокупности, делят пополам. Через полученную точку проводят прямую, параллельную оси абсцисс, до пересечения ее с кумулятой. Абсцисса точки пересечения является медианой.

Мода определяется по гистограмме распределения. Для этого правую вершину модального прямоугольника соединяем с правым верхним углом предыдущего прямоугольника, а левую вершину модального прямоугольника — с левым верхним углом последующего прямоугольника. Абсцисса точки пересечения этих прямых и будет модой распределения (см. рис. 5.4).

Основной характеристикой центра распределения является средняя арифметическая, опирающаяся на всю информацию об изучаемой совокупности единиц. Однако в ряде случаев средняя арифметическая должна быть дополнена и даже заменена модальным значением или медианой. Например, в статистическом контроле качества продукции удобнее пользоваться медианой, а не средней арифметической, так как определение медианы для ранжированного ряда данных не требует специального расчета. Кроме того, она не чувствительна к крайним значениям взятой контрольной пробы. В рядах с открытыми интервалами также целесообразнее пользоваться в качестве характеристики центра распределения модой и медианой. Мода применяется при изучении спроса населения на товары народного потребления (например, на обувь, одежду и т.д.), когда интерес представляет определение модального размера, т.е. размера, пользующегося наибольшим спросом.

Если сформулировать общие правила для выбора средней арифметической, моды или медианы в качестве показателя центра распределения, то можно сказать, что в симметричных рядах все названные показатели равноправны, поскольку $\bar{x} = Me = Mo$, но предпочтение отдается средней арифметической.

Для асимметричных рядов распределения медиана часто является предпочтительной характеристикой центра распределения, поскольку занимает положение между средней арифметической и модой.

5.8. Как рассчитать показатели вариации (колеблемости) признака?

Средняя величина дает обобщающую характеристику всей совокупности изучаемого явления. Однако два ряда распределения, имеющие одинаковую среднюю арифметическую величину, могут значительно отличаться друг от друга по степени колеблемости (вариации) величины изучаемого признака. Если индивидуальные значения признака ряда мало отличаются друг от друга, то средняя арифметическая будет достаточно показательной характеристикой свойств данной совокупности. Если же ряд распределения характеризуется значительным рассеиванием индивидуальных значений признака, то средняя арифметическая будет ненадежной характеристикой этой совокупности и ее практическое применение будет ограничено.

Для измерения вариации признака применяются различные абсолютные и относительные показатели. К абсолютным показателям вариации относятся размах колебаний, среднее линейное отклонение, дисперсия, среднее квадратическое отклонение и квартальное отклонение.

Размах колебаний, или ***размах вариации***, представляет собой разность между максимальным и минимальным значениями признака в изучаемой совокупности:

$$R = x_{max} - x_{min}.$$

Безусловным достоинством этого показателя является простота расчета. Однако размах вариации зависит от величины только крайних значений признака, поэтому область его применения ограничена достаточно однородными совокупностями. В частности, на практике он находит применение в предупредительном контроле качества продукции.

Точнее характеризует вариацию признака показатель, основанный на учете колеблемости всех значений признака. Поскольку средняя арифметическая является обобщающей характеристикой свойств совокупности, большинство показателей вариации основано на рассмотрении отклонений значений признака отдельных единиц совокупности от этой величины. К таким показателям относятся среднее линейное отклонение, дисперсия и среднее квадратическое отклонение, представляющие собой среднюю арифметическую из отклонений индивидуальных значений признака от средней арифметической. Среднее линейное отклонение рассчитывается из отклонений в первой степени, дисперсия и среднее квадратическое — из отклонений во второй степени. Так как алгебраическая сумма отклонений индивидуальных значений признака от

средней арифметической (согласно нулевому свойству) всегда равна нулю, то для расчета среднего линейного отклонения используется арифметическая сумма отклонений, т.е. суммируются абсолютные значения индивидуальных отклонений значений признака независимо от знака.

Формулы для расчета абсолютных показателей вариации приведены в табл. 5.3.

Как видно из табл. 5.3, дисперсия σ^2 представляет собой среднюю из квадратов отклонений значений признака от их средней величины.

Таблица 5.3

Показатели	Формулы расчета	
	по несгруппированным данным	по сгруппированным данным[1]
1. Среднее линейное отклонение ($\bar{d}$)	$\frac{\sum_{i=1}^{n}\lvert x_i - \bar{x}\rvert}{n}$	$\frac{\sum_{i=1}^{k}\lvert x_i - \bar{x}\rvert f_i}{\sum_{i=1}^{k} f_i}$
2. Дисперсия (σ^2)	$\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$	$\frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 f_i}{\sum_{i=1}^{k} f_i}$
2а. Преобразованная формула для расчета дисперсии	$\frac{\sum_{i=1}^{n} x_i^2}{n} - (\bar{x})^2$	$\frac{\sum_{i=1}^{k} x_i^2 f_i}{\sum_{i=1}^{k} f_i} - \left(\frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i}\right)^2$
3. Среднее квадратическое отклонение (σ)	$\sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$	$\sqrt{\frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 f_i}{\sum_{i=1}^{k} f_i}}$

[1] При расчете показателей вариации для интервального ряда распределения вместо x_i подставляются центральные значения интервала x_i'.

Формулу для расчета дисперсии можно преобразовать следующим образом:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^{n}[x_i^2 - 2x_i\bar{x} + (\bar{x})^2]}{n} = \frac{\sum_{i=1}^{n}x_i^2 - 2\bar{x}\sum_{i=1}^{n}x_i + n(\bar{x})^2}{n} =$$

$$= \frac{\sum_{i=1}^{n}x_i^2}{n} - (\bar{x})^2,$$

т.е. дисперсия равна разности средней из квадратов значений признака и квадрата средней арифметической. При пользовании этой формулой исключаются дополнительная процедура расчета отклонений индивидуальных значений признака от $\bar{x}$ и ошибка в расчете, связанная с округлением отклонений ($x_i - \bar{x}$).

Дисперсия обладает рядом свойств, некоторые из них позволяют упростить ее вычисления:

1) дисперсия постоянной величины равна нулю;

2) если все варианты значений признака уменьшить на одно и то же число, то дисперсия не уменьшится;

3) если все варианты значений признака уменьшить в одно и то же число раз (k раз), то дисперсия уменьшится в k^2 раз.

Дисперсия представляется в квадратах единиц, в которых измеряется регистрируемый признак, что делает ее интерпретацию довольно затруднительной. Эту проблему можно преодолеть, рассчитав среднее квадратическое отклонение, которое представляет собой корень квадратный из дисперсии.

Размах вариации, среднее линейное и среднее квадратическое отклонения являются величинами именованными. Они имеют те же единицы измерения, что и индивидуальные значения признака.

Дисперсия и среднее квадратическое отклонение — наиболее широко применяемые показатели вариации. Объясняется это тем, что они входят в большинство теорем теории вероятностей, служащих фундаментом математической статистики. Кроме того, дисперсия может быть разложена на составные элементы, позволяющие оценить влияние различных факторов, обусловливающих вариацию признака. В последующих главах будет показано, как дисперсия используется для построения показателей тесноты корреляционной связи, при оценке результатов выборочных наблюдений и т.д.

Расчет показателей вариации для банков, сгруппированных по размеру прибыли, показан в табл. 5.4.

На основании данных итоговой строки графы 4 определяем среднюю арифметическую $\bar{x} = 121{,}70/20 = 6{,}085$ млн руб.

Для расчета среднего линейного отклонения используем итоги графы 6: $\bar{d} = 17{,}640/20 = 0{,}882$ млн руб.

Дисперсия определяется по данным итоговой строки графы 7:

$\sigma^2 = 23{,}126/20 = 1{,}156$.

Зная дисперсию, рассчитаем среднее квадратическое отклонение:

$$\sigma = \sqrt{1{,}156} = 1{,}075 \text{ млн руб.}$$

Среднее линейное и среднее квадратическое отклонения показывают, на сколько в среднем колеблется величина признака у единиц исследуемой совокупности. Так, в данном примере средняя величина колеблемости размера прибыли составляет: по среднему линейному отклонению 0,882 млн руб., по среднему квадратическому отклонению 1,075 млн руб.

По свойству мажорантности средних величин (см. гл. 4) среднее квадратическое отклонение всегда больше среднего линейного отклонения. Если распределение признака близко к нормальному или симметричному распределению, то между σ и d существует взаимосвязь: $\sigma \approx 1{,}25d$, или $d \approx 0{,}8\sigma$.

Среднее квадратическое отклонение показывает, как расположена основная масса единиц совокупности относительно средней арифметической. В соответствии с теоремой П.Л. Чебышева (1821—1894)

Таблица 5.4

Размер прибыли (в млн руб.)	Число банков	Расчетные показатели				
		$\bar{x}_i$	$\bar{x}_i f_i$	$x_i - \bar{x}$	$\lvert x_i - \bar{x}\rvert f_i$	$(x_i - \bar{x})^2 f_i$
1	2	3	4	5	6	7
3,7—4,6 (–)	2	4,15	8,30	–1,935	3,870	7,489
4,6—5,5	4	5,05	20,20	–1,035	4,140	4,285
5,5—6,4	6	5,95	35,70	–0,135	0,810	0,109
6,4—7,3	5	6,85	34,25	+0,765	3,825	2,926
7,3—8,2	3	7,75	23,25	+1,665	4,995	8,317
Итого	20		121,70		17,640	23,126

Примечание. Знак (–) в первом интервале указывает, что значения признака, совпадающие с верхней границей интервала, включаются в следующий интервал. Это связано с тем, что верхняя граница последнего интервала больше x_{max} в данной совокупности.

можно утверждать, что независимо от формы распределения 75% значений признака попадают в интервал $\bar{x} \pm 2\sigma$, а по крайней мере 89% всех значений попадают в интервал $\bar{x} \pm 3\sigma$.

Вероятности попадания значений признака в интервалы, отклоняющиеся от средней на то или иное число средних квадратических отклонений, для нормального распределения будут рассмотрены в параграфе 5.10.

При вычислении обобщающих статистических характеристик для интервальных рядов распределений действительные значения признака заменяются центральными значениями интервалов, которые более или менее отличаются от средней арифметической из значений, включенных в соответствующий интервал. Так, в первый интервал попали значения 3,7 и 4,3, их средняя арифметическая $\bar{x}$ составит 4,0 млн руб. (3,7 + 4,3/2), а центральное значение первого интервала равно $x_i = 4{,}15$ млн руб. (см. графу 3 табл. 5.4). Во втором интервале средняя прибыль составит 4,95 млн руб. (4,6 + 4,9 + 5,1 + + 5,2/4), в третьем — 5,92 млн руб., в четвертом — 6,84 млн руб. и в пятом — 7,87 млн руб. Таким образом, расхождение в величинах x_i и $\bar{x}_i$ имеет место во всех интервалах, причем наименьшие отклонения ($\bar{x}_i - \bar{x}_i$), соответственно равные +0,03 и +0,01, имеют место в третьем и четвертом интервалах, а наибольшее отклонение, равное +0,15, наблюдается в первом интервале. Можно видеть, что чем больше частота интервала, тем меньше расхождение между величиной средней арифметической из значений признака, включенных в интервал, и центральным значением интервала. Этим объясняется требование достаточно большого числа наблюдений в каждом интервале. Однако, чем далее удалены от центра распределения значения признака, тем больше расхождения между x_i и $\bar{x}_i$.

Если при расчете средней арифметической для достаточно симметричного ряда распределения расхождения между x_i и $\bar{x}_i$ не оказывают существенного влияния на ее отклонение от средней арифметической, рассчитанной по первичным данным, то при расчете дисперсии этот факт приведет к появлению систематической ошибки. В.Ф. Шеппард установил, что ошибка в дисперсии, вызванная применением сгруппированных данных, при расчете составляет $^1/_{12}$ квадрата величины интервала, т.е. скорректированная дисперсия равна

$$\sigma^2 = {}^1/_{12} h^2.$$

Поправка Шеппарда должна применяться при таких условиях, когда распределение:

1) относится к признаку с непрерывным характером вариации;

2) характеризуется тесной близостью с осью абсцисс на концах кривой;

3) построено по большому числу исходных данных (n>500).

Если в качестве показателя центра распределения используется медиана, то для характеристики вариации признака в совокупности можно применить так называемое ***квартильное отклонение Q***. Этот показатель также можно применить вместо размаха вариации, чтобы избежать недостатков, связанных с использованием крайних значений:

$$Q = (Q_3 - Q_1)/2,$$

где Q_1 и Q_2 — соответственно первая и третья квартили распределения.

Квартили — это значения признака в ранжированном ряду распределения, выбранные таким образом, что 25% единиц совокупности будут меньше по величине Q_1; 25% единиц будут заключены между Q_1 и Q_2; 25% — между Q_2 и Q_3 и остальные 25% будут превосходить Q_3. Квартили определяются по формулам, аналогичным приведенной выше формуле для расчета медианы:

$$Q_1 = x_{Q_1} + h \frac{\frac{n+1}{4} - S_{(-1)}}{f_{Q_1}},$$

где x_{Q_1} — нижняя граница интервала, в котором находится первая квартиль;

$S_{(-1)}$ — сумма накопленных частот интервалов, предшествующих интервалу, в котором находится первая квартиль;

f_{Q_1} — частота интервала, в котором находится первая квартиль:

$$Q_3 = x_{Q_3} + h \frac{0{,}75(n+1) - S_{(-1)}}{f_{Q_3}}.$$

(условные обозначения те же, что и для величины Q_1).

По данным табл. 5.4 вычислим:

$$Q_1 = 4{,}6 + 0{,}9 \frac{21/4 - 2}{4} = 5{,}331 \text{ млн руб.};$$

$$Q_2 = 5{,}5 + 0{,}9 \frac{21/4 - 6}{6} = 6{,}175 \text{ млн руб.};$$

$$Q_3 = 6{,}4 + 0{,}9\frac{21 \cdot 3/4 - 12}{5} = 7{,}075 \text{ млн руб.}$$

Используя результаты приведенных расчетов, получим квартильное отклонение:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{7{,}075 - 5{,}331}{2} = 0{,}872 \text{ млн руб.}$$

В симметричных или умеренно асимметричных распределениях $Q = {}^2/_3\sigma$. Так как на квартильное отклонение не влияют отклонения всех значений признака, то его использование следует ограничить случаями, когда определение среднего квадратического отклонения затруднительно или невозможно. В частности, этот показатель может быть рекомендован для рядов распределения с открытыми интервалами, где в качестве характеристики центра распределения использовалась медиана.

5.9. В чем состоит значение относительных показателей вариации?

При сравнении колеблемости различных признаков в одной и той же совокупности или же при сравнении колеблемости одного и того же признака в нескольких совокупностях с различной величиной средней арифметической пользуются ***относительными показателями вариации***. Эти показатели вычисляются как отношение абсолютных показателей вариации к соответствующей характеристике центра распределения — средней арифметической или медиане. Используя в качестве абсолютного показателя вариации размах вариации, среднее линейное отклонение, среднее квадратическое отклонение и квартильное отклонение, получим относительные показатели колеблемости (чаще всего они выражаются в процентах):

Коэффициент осцилляции $K_R = \frac{R}{\bar{x}} \cdot 100\%$.

Относительное линейное отклонение $k_{\bar{d}} = \frac{\bar{d}}{\bar{x}} \cdot 100\%$.

Коэффициент вариации $v = \frac{\sigma}{\bar{x}} \cdot 100\%$.

Относительный показатель квартильной вариации

$$K_Q = \frac{Q}{Me} \cdot 100\% \text{ или } K_Q = \frac{Q_3 - Q_1}{2Q_2} \cdot 100\%.$$

Наиболее часто применяемый показатель относительной колеблемости — ***коэффициент вариации***. Его используют не только для сравнительной оценки вариации разных признаков или в различных совокупностях, но и для характеристики однородности совокупности. Совокупность считается однородной, если коэффициент вариации не превышает 33% (для распределений, близких к нормальному). Для примера, приведенного в табл. 5.4, относительные показатели вариации получились следующими:

$$K_R = \frac{8{,}1 - 3{,}7}{6{,}085} \cdot 100\% = 72{,}31\%;$$

$$k_{\bar{d}} = \frac{0{,}882}{6{,}085} \cdot 100\% = 14{,}5\%;$$

$$v = \frac{1{,}075}{6{,}85} \cdot 100\% = 17{,}6\%;$$

$$K_Q = \frac{0{,}872}{6{,}175} \cdot 100\% = 14{,}12\%.$$

Основываясь на значении коэффициента вариации, можно сделать вывод, что по размеру прибыли совокупность банков является однородной.

Применим коэффициент вариации для сравнительной оценки колеблемости одного признака в разные периоды времени. Например, проводится анализ вариации цен 1 м² муниципального жилья в разных районах Москвы в январе и мае 1995 г. Известно, что в январе среднее квадратическое отклонение составило 0,7603 тыс. руб., а в мае — 0,9754 тыс. руб. Означает ли это, что в мае возросла вариация цен на муниципальное жилье в Москве? Если принять во внимание, что σ рассчитывается по отклонениям от соответствующей средней арифметической, нужно сопоставить средний уровень цен на муниципальное жилье в Москве за январь и май 1995 г. В январе средняя цена 1 м² муниципального жилья составляла 3,88 тыс. руб., а в мае — 5,82 тыс. руб. Следовательно, вывод об увеличении или уменьшении вариации цен на муниципальное жилье в мае 1995 г. может быть сделан после расчета показателя относительной колеблемости уровня признака, т.е. коэффициента вариации.

Расчет коэффициента вариации цены 1 м² муниципального жилья в Москве приводит к следующим результатам:

в январе 1995 г. $v_1 = (0{,}7603/3{,}88) \cdot 100\% = 19{,}59\%$;

в мае 1995 г. $v_2 = (0{,}9754/5{,}82) \cdot 100\% = 16{,}77\%$.

Следовательно, в мае 1995 г. вариация цен 1 м² муниципального жилья в Москве снизилась по сравнению с январем 1995 г., поскольку $v_2 < v_1$.

5.10. На какие составляющие можно разложить общую дисперсию признака?

Если статистическая совокупность разбита на группы по какому-либо признаку, то для оценки влияния различных факторов, определяющих колеблемость индивидуальных значений признака, можно воспользоваться разложением дисперсии на составляющие: межгрупповую и внутригрупповую дисперсии.

Если рассчитать дисперсию признака по всей изучаемой совокупности, т.е. общую дисперсию σ_0^2, то полученный показатель будет характеризовать вариацию признака как результат влияния всех факторов, определяющих индивидуальные различия единиц совокупности. Если же поставить задачу — выделить в составе общей дисперсии ту ее часть, которая обусловлена влиянием какого-либо определенного фактора, то следует разбить изучаемую совокупность на группы, положив в основу группировки интересующий нас фактор. Затем нужно изучить раздельно вариацию признака внутри однородных в отношении данного фактора групп и изменения в величине признака от группы к группе. Выполнение такой группировки позволяет разложить общую дисперсию признака на две дисперсии, одна из которых будет характеризовать часть вариации, обусловленную влиянием фактора, положенного в основу группировки, а вторая — вариацию, происходящую под влиянием прочих факторов (кроме фактора, положенного в основу группировки).

Отклонение индивидуальных значений признака от общей средней $\bar{x}_0$ можно представить так:

$$x_{ij} - \bar{x}_0 = (x_{ij} - \bar{x}_j) + (\bar{x}_j - \bar{x}_0).$$

Вариацию, обусловленную влиянием фактора, положенного в основу группировки, характеризует ***межгрупповая дисперсия*** δ^2, которая является мерой колеблемости частных средних по группам $\bar{x}_j$ вокруг общей средней $\bar{x}_0$ и исчисляется по формуле:

$$\delta^2 = \frac{\sum_{j=1}^{k} (\bar{x}_j - \bar{x}_0)^2 n_j}{\sum_{j=1}^{k} n_j},$$

где k — число групп;

n_j — число единиц в j-й группе;

$\bar{x}_j$ — частная средняя по j-й группе;

$\bar{x}_0$ — общая средняя по совокупности единиц.

Вариацию, обусловленную влиянием прочих факторов, характеризует в каждой группе внутригрупповая дисперсия δ_j^2:

$$\delta_j^2 = \frac{\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)}{n_j}.$$

По совокупности в целом вариация значений признака под влиянием прочих факторов характеризуется средней из внутригрупповых дисперсий:

$$\overline{\sigma^2} = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)}{\sum_{j=1}^{k} n_j}, \quad \text{или} \quad \overline{\sigma^2} = \frac{\sum_{j=1}^{k} \sigma_j^2 n_j}{\sum_{j=1}^{k} n_j}.$$

Между общей дисперсией σ_0^2, средней из внутригрупповых дисперсий $\overline{\sigma^2}$ и межгрупповой δ^2 дисперсией существует соотношение, определяемое правилом сложения дисперсий. Согласно этому правилу, *общая дисперсия равна сумме средней из внутригрупповых и межгрупповой дисперсий:*

$$\sigma_0^2 = \overline{\sigma^2} + \delta^2.$$

Рассмотрим правило сложения дисперсий на следующем примере. По результатам маркетингового обследования туристических фирм, организующих недельные туры в Испанию в различные курортные города, получены следующие данные о вариации стоимости туров в сентябре 1998 г. (табл. 5.5).

Таблица 5.5

Группировка туристических фирм по сегментам рынка в Испании в сентябре 1998 г.

Местоположение курорта	Число туристических фирм n_j	Средняя цена недельного тура (в долл.) $\bar{x}_j$	Дисперсия цен тура в группе σ_j^2
Коста-Брава	7	528,57	2728,04
Коста-дель-Соль	6	588,33	8851,14
Итого	13	556,16	5554,09

Примечание. Цены приведены для гостиниц одного класса.

Вариация цен в обследованной группе туристических фирм, обусловленная различием в местоположении курорта, будет характеризоваться величиной межгрупповой дисперсии.

Средняя цена недельного тура по всем фирмам составляет 556,16 долл.:

$$\bar{x}_0 = \frac{528{,}57 \cdot 7 + 588{,}33 \cdot 6}{13} = 556{,}16 \text{ долл.}$$

Тогда межгрупповая дисперсия будет равна 887,59:

$$\delta^2 = \frac{(528{,}57 - 556{,}16)^2 \cdot 7 + (588{,}33 - 556{,}16)^2 \cdot 6}{13} = 887{,}59.$$

Вариация цен под влиянием всех прочих факторов, кроме местоположения курорта, будет характеризоваться величиной средней из внутригрупповых дисперсий:

$$\overline{\sigma^2} = \frac{2728{,}04 \cdot 7 + 8851{,}14 \cdot 6}{13} = 5554{,}09.$$

Вариация цен на недельные туры в Испанию, обусловленная влиянием всех факторов, формирующих уровень цен в данной группе, состоящей из 13 туристических фирм, определяется величиной общей дисперсии:

$$\sigma_0^2 = 887{,}59 + 5554{,}09 = 6441{,}68.$$

Отсюда можно сделать вывод, что на 13,78% [(887,59/6441,68)× × 100%] дисперсия цен на недельные туры объясняется различиями в местоположении курорта, а на 86,22% [(5554,09/6441,68) × × 100%] — влиянием прочих факторов. Таким образом, преобладающее влияние на вариацию цен недельных туров в Испанию оказывают прочие факторы.

5.11. Как оценить дисперсию альтернативного признака?

Выше был рассмотрен расчет показателей вариации для количественных признаков. Но наряду с вариацией количественных признаков может ставиться задача оценки вариации качественных признаков. При наличии двух взаимоисключающих вариантов значений признака говорят о наличии альтернативной изменчивости качественных признаков. Например, при изучении качества изготовленной продукции можно разделить ее на две группы: годную и бракованную. В таком случае оценивается альтернативный признак. Можно считать, что эквивалентом качественного признака будет переменная, которая принимает значение 1 или 0, причем значение 1 она принимает в том случае, когда обследуемая единица обладает данным признаком, а значение 0, когда не обладает им.

Допустим, общее число единиц совокупности равно n, число единиц, обладающих данным признаком, — f, тогда число единиц, не обладающих данным признаком, будет равно $(n - f)$. Учитывая изложенное, построим ряд распределения по качественному признаку:

Значение переменной	Частота повторений
1	f
0	$n - f$
Итого	n

Средняя арифметическая такого ряда равна:

$$\bar{x} = \frac{1 \cdot f + O \cdot (n - f)}{n} = \frac{f}{n},$$

т.е. равна относительной частоте (частости), которую можно обозначить через p, тогда $\bar{x} = p$.

Таким образом, доля единиц, обладающих данным признаком, равна p; соответственно доля единиц, не обладающих данным признаком, равна q; $p + q = 1$. Тогда дисперсия альтернативного признака определяется по формуле:

$$\sigma^2 = \frac{(1 - p)^2 p + (0 - p)^2 q}{p + q} = \frac{q^2 p + p^2 q}{p + q} = pq,$$

а среднее квадратическое отклонение альтернативного признака:

$$\sigma = \sqrt{pq}.$$

Например, в результате контроля качества при приемке из 1000 готовых изделий 20 оказались бракованными. Применяя вышеуказанную символику, строим ряд распределения (1 соответствует бракованным изделиям, а 0 — годной продукции):

Значение переменной	Частота повторений
1	20
0	980
Итого	1000

Доля брака по данным примера равна 2% (20/1000 × 100%). Тогда величина дисперсии составит 0,0196 ($\sigma_p^2 = 0{,}02 \times 0{,}98$), а среднее квадратическое отклонение $\sigma_p = 0{,}14$.

5.12. Что характеризуют показатели дифференциации?

Анализ вариации в рядах распределения целесообразно дополнить ***показателями дифференциации***. По первичным данным может быть рассчитан так называемый *коэффициент фондовой дифференциации* $K_ф$, который представляет собой соотношение двух средних, полученных из 10% наибольших и наименьших значений признака. Так, по данным о размере прибыли по 20 коммерческим банкам (см. табл. 5.2) находим, что два банка (10% от общего числа) с наименьшим размером прибыли 3,7 млн и 4,3 млн руб. имеют $\bar{x}_{наим} = 4{,}0$ млн руб. Для двух банков с наибольшими значениями 7,9 млн и 8,1 млн руб. средняя арифметическая $\bar{x}_{наиб}$ составит 8,0 млн руб. Таким образом, коэффициент фондовой дифференциации будет равен 2,0: $K_ф = \bar{x}_{наиб} / \bar{x}_{наим} = 8{,}0/4{,}0 = 2{,}0$.

Это означает, что размер прибыли у 10% банков с наивысшими доходами в два раза превышает размер прибыли 10% коммерческих банков с наименьшими доходами[5].

[5] Этот пример имеет чисто методический характер, так как рассчитывать коэффициенты дифференциации при малом объеме совокупности нецелесообразно.

Если представлены сгруппированные данные, то для характеристики дифференциации можно воспользоваться соотношением десятой и первой децили (децили делят все число единиц в совокупности на 10 равных частей). Для определения децилей используются формулы, аналогичные тем, что приведены выше для расчета квартилей. Общая схема их расчета такова:

1) определяется номер для первой децили и для девятой децили N_{D_1}:

$$N_{D_1} = \frac{n+1}{10}; \quad N_{D_9} = \frac{9(n+1)}{10};$$

2) устанавливается интервал, где должны будут находиться децили;

3) рассчитывается значение децилей при предположении равномерного наращения величины интервала на каждую единицу частоты (частости);

4) определяется *коэффициент дециальной дифференциации*:

$$K_D = D_9/D_1.$$

Покажем расчет этого коэффициента на основе данных о распределении населения России по размеру среднедушевого денежного дохода в июне и сентябре 1998 г. (табл. 5.6).

По данным за июнь первая дециль будет находиться в первом интервале — до 400 руб., а девятая — в интервале от 1200 до 1600 руб. Их значение составляет соответственно 308,11 руб. и 1658,33 руб.

Таблица 5.6

Распределение населения Российской Федерации по размеру среднедушевого дохода в июне и сентябре 1998 г.

Ежемесячный среднедушевой доход (в руб.)	Численность населения (в % к итогу)		Накопленная частость (в %)	
	июнь	сентябрь	июнь	сентябрь
До 400	18,5	15,2	18,5	15,2
400—600	21,4	19,6	39,9	34,8
600—800	18,0	17,7	57,9	52,5
800—1000	13,1	13,6	71,0	66,1
1000—1200	9,0	9,8	80,0	75,9
1200—1600	10,3	11,7	90,3	87,6
1600—2000	4,8	5,8	95,1	93,4
Свыше 2000	4,9	6,6	100,0	100,0
Итого	100,0	100,0		

$$D_1 = 200 + 200\frac{10}{18,5} = 308,11 \text{ руб.};$$

$$D_9 = 1200 + 400\frac{90 - 80,0}{10,3} = 1588,35 \text{ руб.}$$

В распределении по среднедушевому доходу за сентябрь первая дециль также будет находиться в первом интервале, а девятая дециль – в интервале 1600-2000 руб.:

$$D_1 = 200 + 200\frac{10}{15,2} = 331,57 \text{ руб.};$$

$$D_9 = 1600 + 400\frac{90 - 87,6}{5,8} = 1765,51 \text{ руб.};$$

в июне 1998 г. $K_D = 1588,35/308,11 = 5,16$;
в сентябре 1998 г. $K_D = 1765,51/331,57 = 5,32$.

Таким образом, наименьший уровень среднедушевого денежного дохода 10% наиболее обеспеченного населения и наивысший уровень среднедушевого денежного дохода 10% наименее обеспеченного населения в июне 1998 г. различались в 5,16 раза, а в сентябре — в 5,32 раза.

5.13. Что характеризуют показатели формы распределения?

В практике статистического исследования приходится встречаться с самыми различными распределениями. Однородные совокупности характеризуются, как правило, одновершинными распределениями. Многовершинность свидетельствует о неоднородности изучаемой совокупности. Появление двух и более вершин говорит о необходимости перегруппировки данных с целью выделения более однородных групп. Выяснение общего характера распределения предполагает оценку степени его однородности, а также вычисление показателей асимметрии и эксцесса. *Симметричным является распределение, в котором частоты любых двух вариантов, равно отстоящих в обе стороны от центра распределения, равны между собой.* Для симметричных распределений имеет место равенство средней арифметической, моды и медианы. В связи с этим *простейший показатель асимметрии основан на соотношении показателей центра распределения: чем больше разница между средними* $(\bar{x} - Mo)$, *тем больше асимметрия ряда.*

Для сравнительного анализа степени асимметрии нескольких распределений рассчитывают относительный показатель *As:*

$$As = \frac{(\bar{x} - Mo)}{\sigma}.$$

Величина показателя асимметрии *As* может быть положительной и отрицательной. Положительная величина показателя асимметрии указывает на наличие правосторонней асимметрии (правая ветвь относительно максимальной ординаты вытянута больше, чем левая; рис. 5.7а). При правосторонней асимметрии между показателями центра распределения существует соотношение $Mo < Me < \bar{x}$. Отрицательный знак показателя асимметрии свидетельствует о наличии левосторонней асимметрии (рис. 5.7б). Между показателями центра распределения в этом случае существует такое соотношение:

$Mo > Me > \bar{x}$.

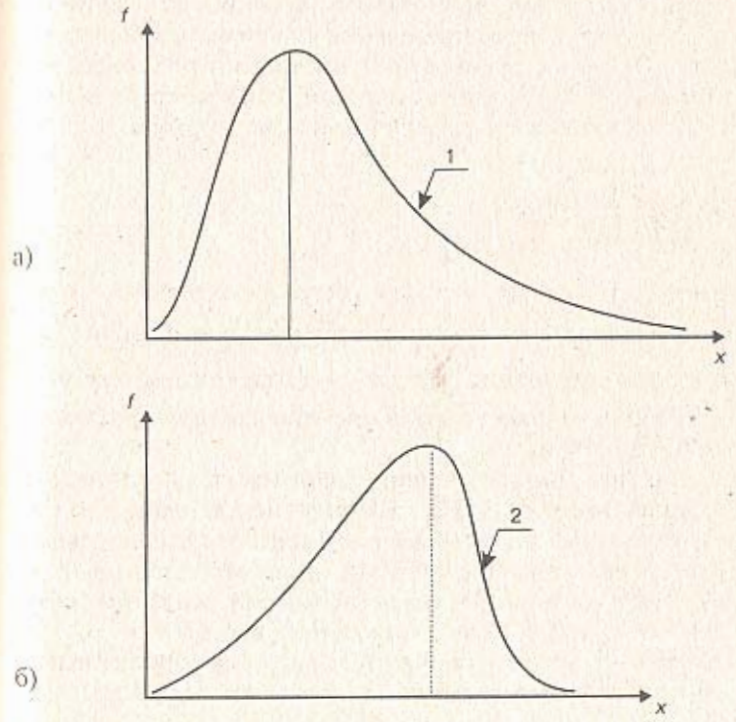


Рис. 5.7. Асимметричные ряды распределения:
(а) с правосторонней асимметрией;
(б) с левосторонней асимметрией.

Коэффициент асимметрии может быть рассчитан и с помощью квартилей:

$$K_{As} = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}.$$

Наиболее точным и распространенным является показатель, основанный на определении центрального момента третьего порядка (в симметричном распределении его величина равна нулю):

$$As = \frac{\mu_3}{\sigma^3},$$

где $\mu_3 = \dfrac{\sum_{i=1}^{k}(x_i - \bar{x})^3 f_i}{\sum_{i=1}^{k} f_i}$.

Применение этого показателя дает возможность не только определить степень асимметрии, но и ответить на вопрос о наличии или отсутствии асимметрии в распределении признака в генеральной совокупности. Оценка степени существенности этого показателя дается с помощью средней квадратической ошибки, которая зависит от объема наблюдений и рассчитывается по формуле:

$$\sigma_{As} = \sqrt{\frac{6(n-1)}{(n+1)(n+3)}}.$$

Если отношение $|As| / \sigma_{As} > 3$, асимметрия существенна и распределение признака в генеральной совокупности не является симметричным. Если отношение $|As| / \sigma_{As} < 3$, то асимметрия несущественна, ее наличие может быть объяснено влиянием различных случайных обстоятельств.

Для симметричных распределений рассчитывается показатель ***эксцесса*** (островершинности). На рис. 5.8 представлены два распределения: одно — островершинное (величина эксцесса положительная), другое — плосковершинное (величина эксцесса отрицательная). *Эксцесс представляет собой выпад вершины эмпирического распределения вверх или вниз от вершины кривой нормального распределения.*

Наиболее точным является показатель эксцесса, основанный на использовании центрального момента четвертого порядка:

$$Ex = \frac{\mu_4}{\sigma^4} - 3.$$

В нормальном распределении отношение $\dfrac{\mu_4}{\sigma^4} = 3$.

Средняя квадратическая ошибка эксцесса рассчитывается по формуле:

$$\sigma_{Ex} = \sqrt{\frac{24n(n-2)(n-3)}{(n-1)^2(n+3)(n+5)}},$$

где n — число наблюдений.

Оценка существенности показателей асимметрии и эксцесса позволяет сделать вывод о том, можно ли отнести данное эмпирическое распределение к типу кривых нормального распределения.

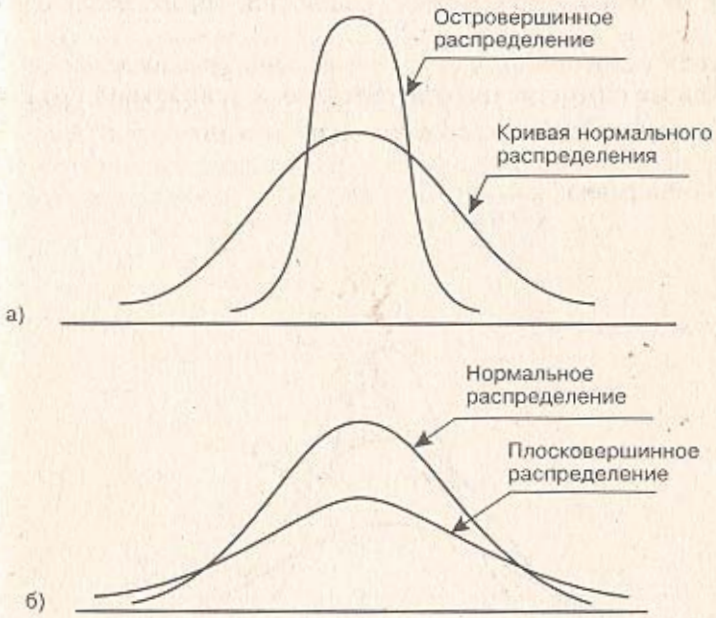


Рис. 5.8. Ряды распределения с положительным (а) и отрицательным (б) эксцессом

5.14. Каковы особенности кривой нормального распределения?

Если непрерывная случайная величина имеет плотность распределения:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}},$$

то она подчиняется закону нормального распределения. Для построения кривой нормального распределения надо знать два параметра — $\bar{x}$ и σ.

Если средняя арифметическая не меняется, но растет величина среднего квадратического отклонения, распределение имеет более плосковершинный характер (рис. 5.9).

На рис. 5.10 приведено «семейство» кривых нормального распределения с одной и той же величиной среднего квадратического отклонения, но с разными средними ($\bar{x}_1 < \bar{x}_2 < \bar{x}_3$). В этом случае кривая, не меняя своей формы, сдвигается вправо вдоль оси абсцисс.

Укажем особенности кривой нормального распределения.

1. Кривая симметрична относительно максимальной ординаты. Максимальная ордината соответствует значению $\bar{x} = Mo = Me$, ее величина равна $\frac{1}{\sqrt{2\pi}\sigma}$.

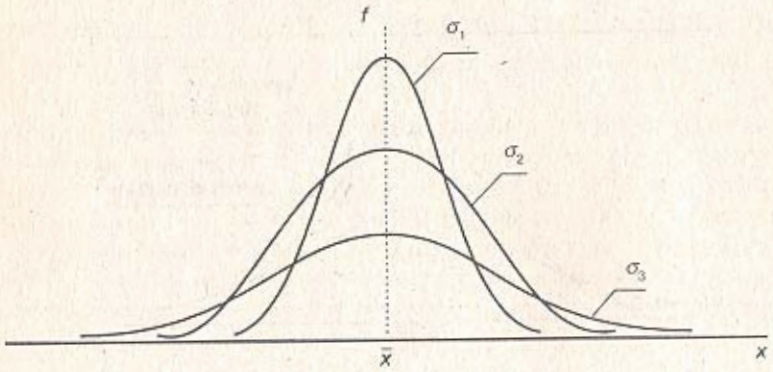


Рис. 5.9. Кривые нормального распределения с одинаковой $\bar{x}$, но разными σ ($\sigma_1<\sigma_2<\sigma_3$)

2. Кривая асимптотически приближается к оси абсцисс, продолжаясь в обе стороны до бесконечности. Следовательно, чем больше значения отклоняются от $\bar{x}$, тем реже они встречаются. Одинаковые по абсолютному значению, но противоположные по знаку отклонения значений переменной x от $\bar{x}$ равновероятны.

3. Кривая имеет две точки перегиба, находящиеся на расстоянии $\pm\sigma$ от $\bar{x}$.

4. При $\bar{x}$ = const с увеличением σ кривая становится более пологой. При σ = const с изменением $\bar{x}$ кривая не меняет свою форму, а лишь сдвигается вправо или влево по оси абсцисс.

5. В промежутке $\bar{x} \pm\sigma$ находится 68,3% всех значений признака. В промежутке $\bar{x} \pm 2\sigma$ находится 95,4% всех значений признака. В промежутке $\bar{x} \pm 3\sigma$ находится 99,7% всех значений признака (рис. 5.11).

Эти свойства кривой нормального распределения весьма полезны. Например, известна средняя величина в ряду распределения, подчиняющемся нормальному распределению, которая равна 150, а среднее квадратическое отклонение равно 5. Тогда нам будет известно, что 68,3% всех значений признака в исследуемой совокупности будут иметь значения между 145 и 155 (150±5), а 99,7% всех значений признака (т.е. почти у всех наблюдаемых единиц) будут находиться между 135 и 165 (150±3×5).

Нормальное распределение возможно в том случае, когда на величину признака влияет большое число случайных причин. Действие этих причин независимо, и ни одна из причин не имеет преобладающего влияния на другие.

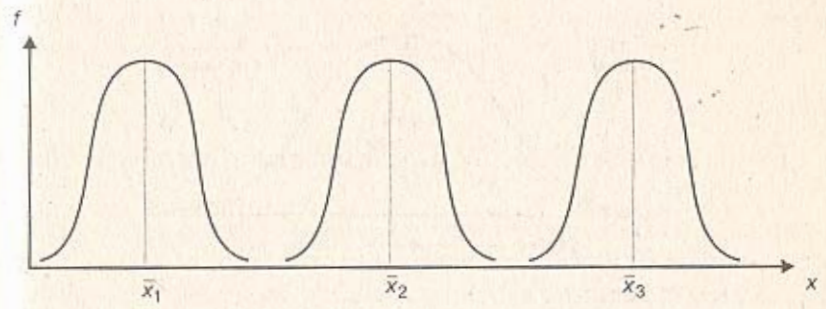


Рис. 5.10. Кривые нормального распределения с одинаковым σ, но разными $\bar{x}$ ($\bar{x}_1<\bar{x}_2<\bar{x}_3$)

Для удобства вычислений вероятностей случайные величины нормируются, а затем используются заранее табулированные значения плотности функции распределения нормированной случайной величины. Если обозначим $\frac{x-\bar{x}}{\sigma}$ через t, то величину $\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$ назовем нормированной функцией; эта функция табулирована. Для нормированной случайной величины математическое ожидание равно нулю, а дисперсия равна единице.

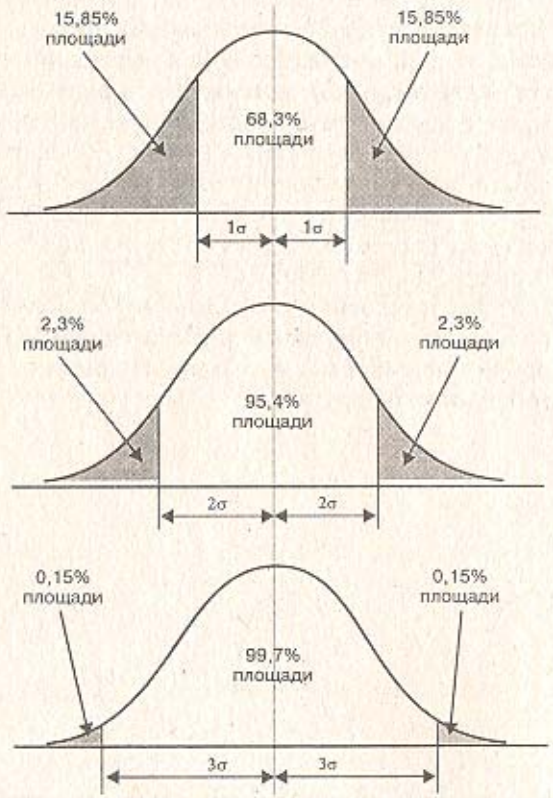


Рис. 5.11. Соотношение площади под кривой нормального распределения в зависимости от расстояния от средней арифметической

Определенный интеграл вида $F(t)=\frac{1}{\sqrt{2\pi}}\int_0^t e^{\frac{u^2}{2}}$ носит название нормированной функции Лапласа и характеризует площадь под кривой в промежутке от нуля до t (см. Приложение 2).

Чтобы оценить вероятность попадания в интервал от $-\infty$ до x, рассчитываем $F(x) = {}^1/_2 + F(t)$. Для определения вероятности попадания нормально распределенной случайной величины x в заданный интервал $(x_1; x_2)$ находим разность $F(x_2) - F(x_1)$:

$$P(x_1 \le x \le x_2) = F(x_2) - F(x_1) = \left[\frac{1}{2} + F(t_2)\right] - \left[\frac{1}{2} + F(t_1)\right] =$$

$$= F(t_2) - F(t_1),$$

где $t_1 = \frac{x_1 - \bar{x}}{\sigma}$; $t_2 = \frac{x_2 - \bar{x}}{\sigma}$.

Например, текущая цена акций примерно подчиняется нормальному закону со средней 153 руб. и средним квадратическим отклонением 1,2 руб. Определим вероятность того, что цена акций будет находиться между 150 руб. и 155 руб.

Значениям цены акций 150 и 155 руб. будут соответствовать нормированные отклонения:

$t_1 = (150 - 159)/1{,}2 = -2{,}5$ и $t_2 = (155 - 153)/1{,}2 = 1{,}67$.

Значения нормированной функции, соответствующие данным значениям t_1 и t_2, находятся по Приложению 2:

$F(t_1 = -2{,}5) = -0{,}4938$; $[F(-t) = -F(t)]$;

$F(t_2 = 1{,}67) = 0{,}4118$.

Разность $F(t_2) - F(t_1) = 0{,}4118 - (-0{,}4938) = 0{,}9056$.

Таким образом, вероятность того, что цена акции будет находиться в пределах от 150 до 155 руб., равна 90,56%.

5.15. Что такое вариационный ряд и группировка?

Знакомство со свойствами и назначением основных показателей, характеризующих вариационные ряды, а также с особенностями кривой нормального распределения позволяет более подробно остановиться на принципах группировки исходных данных. Как уже указывалось, в статистических исследованиях должно обеспечиваться единство анализа качественной и количественной сторон изучае-

мых явлений. Одно из важнейших требований теории статистики к научно обоснованному анализу – соблюдение ***принципа однородности*** исследуемых единиц. *Необходимое условие однородности— однотипность явлений,* устанавливаемая уже при первых шагах статистического исследования. Затем в рамках выполнимости необходимого условия проверяется и количественная однородность как достаточное условие. Обычно при построении вариационного ряда речь идет не о выделении типов, а о количественной характеристике вариации. Однако в процессе такого анализа может возникнуть проблема выделения в составе исследуемой совокупности достаточно различающихся между собой групп. При построении рядов распределений важное значение имеет решение вопроса о выборе *величины интервала,* что связано с характеристикой формы распределения и точностью вычисляемых оценок. Так, Р. Фишер подчеркивает значение выбора интервала для характеристики формы распределения: «В тех весьма частых случаях, когда переменная изменяется непрерывно, так что возможны все промежуточные значения, выбор размера интервалов и их границ произволен, а это может привести к заметным различиям во внешнем виде диаграмм»[6].

При расчете эмпирических характеристик по сгруппированным данным в расчетные формулы вместо фактических значений подставляются центральные значения интервалов. И здесь естественно поставить вопрос, в какой мере оценки средней и среднего квадратического отклонения, полученные по сгруппированным данным, близки к таким же оценкам, вычисленным без группировки. Вычисления будут точными, если средняя арифметическая всех значений, попавших в соответствующий интервал, совпадает с его центральным значением. Если отклонения средних арифметических интервалов от центральных значений не имеют систематического характера, то будет достигнута достаточная точность. Таким образом, ширина интервала должна быть такой, чтобы центральное значение интервала давало достаточно близкое представление о средней величине всех тех индивидуальных значений, которые входят в соответствующий интервал. Кроме того, весьма важно, чтобы число наблюдений в интервале не было слишком малым, а соответственно число групп чрезмерно большим. В общем виде можно сказать, что расчет величины интервала должен основываться на учете степени вариации признака и численности единиц исследуемой совокупности.

В условиях машинной обработки информации речь должна идти о тех или иных стандартных процедурах группировки по количественным признакам. Один из вариантов такого стандарта приведен в параграфе 5.2, где для определения величины интервала предложена формула Стэрджесса.

[6] Фишер Р. Статистические методы для исследователей. М.: Статистика, 1958. С. 35.

В дополнение к уже рассмотренной процедуре при построении рядов распределения можно воспользоваться свойствами кривой нормального распределения (см. параграф 5.14). Учитывая, что в пределах $\bar{x} \pm 3\sigma$ должно находиться 99,7% всех вариантов значений признака, можно предложить следующие интервалы группировки:

1) до $\bar{x} - 3\sigma$; 2) $\bar{x} - 3\sigma \div \bar{x} - 2\sigma$; 3) $\bar{x} - 2\sigma \div \bar{x} - \sigma$; 4) $\bar{x} - \sigma \div \bar{x}$; 5) $\bar{x} \div \bar{x} + \sigma$; 6) $\bar{x} + \sigma \div \bar{x} + 2\sigma$; 7) $\bar{x} + 2\sigma \div \bar{x} + 3\sigma$; 8) $\bar{x} + 3\sigma$ и более.

Такой вариант группировки может использоваться в совокупностях достаточно большого объема, так как число групп равно 8.

Для асимметричных распределений может быть предложен вариант группировки с использованием средних величин. Общая средняя величина $\bar{x}_0$, рассчитанная по всей совокупности единиц, позволяет выделить две группы с уровнем признака ниже и выше средней. По этим группам рассчитываются средние $\bar{x}_1$ и $\bar{x}_2$, на основе которых в свою очередь можно аналогичным образом сформировать подгруппы. Таким образом, получим следующие интервалы группировки: 1) менее $\bar{x}_1$; 2) $\bar{x}_1 \div \bar{x}_0$; 3) $\bar{x}_0 \div \bar{x}_2$; 4) $\bar{x}_2$ и более.

В совокупностях, достаточно больших по объему, может быть продолжено деление сформированных групп на подгруппы и получено большее число групп.

Покажем рассмотренный вариант группировки на примере, приведенном в табл. 5.2. По несгруппированным данным $\bar{x}_0 = 6{,}055$ млн руб. Тогда в первую группу (ниже средней арифметической) попадут банки, имеющие размер прибыли менее 6,055 млн руб., а именно 3,7; 4,3; 5,6; 5,1; 4,6; 5,7; 5,9; 5,2; 5,8; 4,9, всего 10 банков. Остальные десять образуют группу с размером прибыли выше средней арифметической (в млн руб.): 6,7; 8,1; 6,4; 6,2; 6,3; 7,2; 7,9; 7,6; 7,0; 6,9.

Вычислим средние арифметические по этим группам:

$$\bar{x}_1 = 50{,}8/10 = 5{,}08 \text{ млн руб.};$$

$$\bar{x}_2 = 70{,}3/10 = 7{,}03 \text{ млн руб.}$$

Учитывая, что $x_{min} = 3{,}7$ млн руб., а $x_{max} = 8{,}1$ млн руб., получим группировку, показанную в табл. 5.7.

Поскольку представленный в табл. 5.7 ряд распределения имеет неравные интервалы, в качестве веса для определения $\bar{x}$ следует воспользоваться показателями плотности распределения (см. графу 3):

Таблица 5.7

Группы банков по размеру прибыли (в млн руб.)	Число банков f_i	Плотность распределения f_i / h_i	Центральные значения интервала f_i / h	$x_i \cdot (f_i / h_i)$
1	2	3	4	5
3,7—5,1 (+)	5	3,57	4,40	15,71
5,2—5,9	5	7,14	5,55	39,63
6,0—6,9	5	5,56	6,45	35,86
7,0—8,1	5	4,55	7,55	34,35
Итого	20	20,82		125,55

Примечание: Знак (+), поставленный у верхней границы первого интервала, означает, что значения признака, совпадающие с верхней границей соответствующего интервала, включаются в этот интервал.

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i \cdot \frac{f_i}{h_i}}{\sum_{i=1}^{k} \frac{f_i}{h_i}} = \frac{125,55}{20,82} = 6,030 \text{ млн руб.}$$

Расчет средней арифметической по приведенной в этой таблице группировке дает следующий результат: $\bar{x} = 6,030$ млн руб. Средняя арифметическая, рассчитанная по группировке, приведенной в табл. 5.2, равна 6,085 млн руб.

Таким образом, величина абсолютной ошибки в расчете средней арифметической ($\Delta\bar{x}$) по различным группировкам составила[7]:

$\Delta\bar{x} = 6,085 - 6,055 = 0,030$ млн руб., или (в относительном выражении) 0,5% (0,030/6,055 × 100%) (расчет по табл. 5.2);

$\Delta\bar{x} = 6,030 - 6,055 = -0,025$ млн руб.,
или 0,4% (расчет по табл. 5.7).

Приведенные расчеты позволяют сделать вывод о достаточной точности результатов вычислений средних арифметических значений по обоим вариантам группировки.

В табл. 5.7 мы получили равномерное распределение общего числа наблюдений по группам — в каждую группу попало пять банков. Однако такое распределение единиц по группам совсем не обязательно при использовании приведенного выше алгоритма группировки. Принцип равной наполненности групп по числу наблюдений может быть обеспечен путем использования для определения границ интервалов квартилей (в каждую группу включают 25% общего числа единиц совокупности) или при большом числе наблюдений — децилей (в каждую группу включают 10% от общего числа наблюдений). Использование квартилей для группировки 20 коммерческих банков по размеру прибыли приводит к такому же результату, который уже получен в табл. 5.7.

[7] В качестве базы для сравнений используется величина средней арифметической по несгруппированным данным $\bar{x} = 6,055$ млн руб.

Глава 6
ВЫБОРОЧНОЕ НАБЛЮДЕНИЕ

6.1. В чем состоят преимущества выборочного наблюдения в сравнении с другими видами статистических наблюдений?

Выборочное наблюдение представляет собой один из наиболее широко применяемых видов несплошного наблюдения. При проведении выборочного наблюдения, как и всякого несплошного наблюдения, обследуются не все единицы изучаемого объекта, или, иными словами, обследуются не все единицы генеральной совокупности, а лишь некоторая, так или иначе отобранная часть этих единиц. Однако наблюдение организовано таким образом, что эта часть отобранных единиц в уменьшенном масштабе представляет всю совокупность, т. е. является репрезентативной. Часть единиц генеральной совокупности, подлежащей непосредственному наблюдению, называют выборочной совокупностью.

Система правил отбора единиц и способов характеристики изучаемой совокупности исследуемых единиц составляет содержание выборочного метода.

В последние годы выборочные обследования стали широко применяться в работе органов государственной статистики. Крупные и средние предприятия охватываются сплошным наблюдением, а наблюдение за деятельностью малых предприятий производится с помощью выборочных обследований. В ряде случаев выборочные наблюдения применяются в сочетании со сплошными переписями и учетами. Например, программа Всероссийской переписи населения 1999 г. содержит как вопросы сплошного наблюдения, относящиеся ко всему населению, так и вопросы выборочного наблюдения 25% населения для характеристики основного занятия занимаемого положения, места работы, а также вопросы 5%-ного выборочного обследования с целью изучения брачности и рождаемости.

Применение выборочного наблюдения взамен сплошного дает возможность лучше организовать наблюдение, обеспечивает быстроту проведения наблюдения, приводит к экономии средств и затрат труда на получение и обработку информации. Объективную гарантию репрезентативности полученной выборочной совокупности дает применение соответствующих научно обоснованных способов отбора подлежащих обследованию единиц. В процессе формирования выборочной совокупности должен быть обеспечен строго объективный подход к отбору единиц. Нарушение этого принципа, когда наблюдению подвергаются единицы, отобранные на основании субъективного мнения исследователя, приводит к тому, что результаты такого наблюдения относятся не ко всей генеральной совокупности, а только к той ее части, которая была подвергнута наблюдению.

Наиболее просто случайный отбор единиц можно организовать для совокупностей, учитываемых по состоянию на данный момент и включающих счетное количество единиц. В таких случаях есть возможность заранее составить пронумерованный список всех единиц генеральной совокупности. Отбор из списка единиц может быть произведен с помощью таблиц случайных чисел (см. Приложение 3) или датчика случайных чисел. В соответствии с определенным объемом выборки последовательно отбирают n единиц, номера которых были включены в выборку.

Таким образом, при простой случайной выборке отбор производится из всей массы единиц генеральной совокупности без предварительного разделения ее на какие-либо группы и единица отбора совпадает с единицей наблюдения.

Для отбора единиц можно применить два способа:

1) отбор по «схеме возвращенного шара», обычно называемый повторным отбором. При повторном отборе вероятность попадания каждой отдельной единицы в выборку остается постоянной, так как после отбора какой-либо единицы она снова возвращается в совокупность и опять может быть выбранной;
2) отбор по «схеме невозвращаемого шара», называемый бесповторным отбором. В этом случае каждая отобранная единица не возвращается обратно, и вероятность попадания отдельных единиц в выборку все время изменяется (для оставшихся единиц она возрастает).

В сравнении с другими видами несплошных наблюдений преимущество выборочного наблюдения заключается в том, что по результатам этого наблюдения можно оценить искомые параметры генеральной совокупности. Между характеристиками выборочной совокупности и искомыми характеристиками (параметрами) генеральной совокупности, как правило, существует некоторое расхождение, которое называют ошибкой. Общая величина возможной ошибки выборочной характеристики слагается из ошибок двоякого рода: ошибки регистрации и ошибки репрезентативности.

Ошибки регистрации свойственны любому статистическому наблюдению вообще, их появление может быть вызвано несовершен-

ством измерительных приборов, недостаточной квалификацией наблюдателя, неточностью подсчетов и т.п. Можно предположить, что по сравнению со сплошными наблюдениями опасность возникновения ошибок регистрации при проведении выборочных наблюдений должна быть меньше, так как выборочные наблюдения проводятся с участием более квалифицированных работников и, следовательно, более тщательно. Значительно уменьшается при выборочном наблюдении и опасность преднамеренных искажений данных, так как специально подобранные и обученные наблюдатели в них не заинтересованы.

Возникновение ***случайных ошибок*** репрезентативности объясняется недостаточно равномерным представлением в выборочной совокупности различных категорий единиц генеральной совокупности, в силу чего распределение отобранной совокупности единиц не вполне точно воспроизводит распределение единиц генеральной совокупности. Известен одни хрестоматийный пример некорректного применения выборочного метода во время президентских выборов в США в 1936 г. (приведенный У. Дж. Рейхманом[8]). Один из журналов по адресам телефонной книги и подписчиков журнала отобрал 10 млн человек, которым были разосланы анкеты. На анкеты ответило свыше 2 млн человек, т.е. более 20% отобранных лиц. Ответы опрошенных указывали на то, что Франклин Д. Рузвельт потерпит поражение, тогда как в действительности он был избран большинством голосов, причем самым представительным в истории выборов в США. Причиной провала попытки дать точное предсказание результатов выборов было то, что выборка не была представительной для всей совокупности американских избирателей. В те времена телефон был предметом роскоши и его имело больше республиканцев, чем демократов. Поэтому, пользуясь телефонной книгой, организаторы обследования, по существу, отбирали избирателей, принадлежавших преимущественно к партии республиканцев.

Во время тех же выборов институтом Гэллапа было собрано четыре тысячи анкет (против двух миллионов, собранных журналом), однако он точно предсказал долю голосовавших за Франклина Д. Рузвельта.

Таким образом, размер отобранной совокупности не может служить гарантией правильной оценки параметров в генеральной совокупности, если нарушены принципы научной организации выборочного наблюдения.

Определение ***возможной*** и ***фактически допущенной ошибки*** выборки имеет важное значение при применении выборочного метода. Величина ошибки характеризует степень надежности результатов выборки; знание этой величины необходимо при оценке параметров генеральной совокупности. Оценки возможной величины и состава ошибок репрезентативности ложатся в основу планирования проектируемого выборочного наблюдения.

[8] Рейхман У.Дж. Применение статистики. М.: Статистика, 1969. С. 222.

Величина случайной ошибки репрезентативности зависит:

от принятого способа формирования выборочной совокупности. Выбор последнего связан с решением вопросов о единице отбора, способе отбора единиц, способе размещения всего объема отбираемых единиц по различным группам генеральной совокупности;

от объема выборки;

от степени колеблемости изучаемого признака в генеральной совокупности.

Для каждого конкретного выборочного наблюдения величина ошибки репрезентативности может быть определена по соответствующим формулам.

6.2. Каковы теоретические основы выборочного метода?

Развитие современной теории выборочного наблюдения началось с простой случайной выборки. Лежащие в основе простой случайной выборки понятия и категории являются исходными при разработке других форм выборочного наблюдения.

В математической теории выборочного наблюдения доказывается, что с увеличением объема выборки вероятность появления больших ошибок и пределы максимально возможной ошибки уменьшаются (чем больше обследуется единиц, тем меньше будет величина расхождений выборочных и генеральных характеристик).

Принципиальную возможность определения генеральной средней по данным простой случайной выборки доказывает теорема П.Л. Чебышева. В приложении к выборочному методу неравенство П.Л. Чебышева может быть сформулировано так: при неограниченном увеличении числа независимых наблюдений ($n \to \infty$) в генеральной совокупности с ограниченной дисперсией с вероятностью, сколь угодно близкой к единице, можно ожидать, что отклонение выборочной средней от генеральной средней будет сколь угодно мало:

$$P\left(|\bar{x} - \bar{x}| \le \varepsilon\right) \to 1 \text{ при } n \to \infty,$$

где P — вероятность неравенства, стоящего в скобках;

ε — любое, сколь угодно малое положительное число;

$\tilde{x}$ — выборочная средняя;

$\bar{x}$ — генеральная средняя;

Однако, пользуясь этой теоремой, мы не можем указать вероятность появления ошибок определенной величины.

На этот вопрос отвечает центральная предельная теорема А. М. Ляпунова, доказанная в 1901 г. Согласно этой теореме, при достаточно большом числе независимых наблюдений в генеральной совокупности с конечной средней и ограниченной дисперсией вероятность того, что расхождение между выборочной и генеральной средней $|\tilde{x} - \bar{x}|$ не превзойдет по абсолютной величине некоторую величину, равна интегралу Лапласа. Можно записать сказанное таким образом:

$$P\left(|\tilde{x} - \bar{x}| \le t\mu\right) \to F(t),$$

где $F(t)$ — нормированная функция Лапласа:

$$F(t) = \frac{1}{\sqrt{2\pi}} \int_{-t}^{+t} e^{-\frac{t^2}{2}} dt = \frac{2}{\sqrt{2\pi}} \int_{0}^{t} e^{-\frac{t^2}{2}} dt.$$

Величина μ есть средняя квадратическая (стандартная) ошибка выборки.

Из теоремы А.М. Ляпунова непосредственно следует, что при достаточно большом числе независимых наблюдений распределение выборочных средних (следовательно, и их отклонений от генеральной средней) приближенно нормально.

Частным случаем теоремы П.Л. Чебышева является теорема Я. Бернулли:

$$p\left[(w - p) < t\sqrt{\frac{p(1-p)}{n}}\right] > 1 - \frac{1}{t^2},$$

где w и p — доля признака собственно в выборочной и генеральной совокупности.

Таким образом, практика применения выборочного метода опирается на ряд закономерностей простой случайной выборки:

1. Из возможных результатов простой повторной выборки наиболее вероятны такие, при которых величина выборочной средней будет близка к величине генеральной средней, и, следовательно, разность будет близка к нулю.

2. В отдельных выборках значения выборочных средних не будут точно совпадать с величиной генеральной средней, но если рассчитать среднюю из всех возможных значений выборочной средней (математическое ожидание выборочной средней), то величина этой

средней будет совпадать с величиной генеральной средней. Другими словами, при простой случайной выборке средняя арифметическая является несмещенной оценкой генеральной средней.

3. Чем больше величина случайной ошибки выборки, тем менее вероятно появление такой ошибки. Всегда можно указать с заданной вероятностью предел расхождений между выборочной и генеральной средней.

6.3. Как определить стандартную и предельную ошибку простой случайной выборки?

В математической статистике доказывается, что величина средней квадратической (стандартной) ошибки простой случайной выборки может быть определена по формуле:

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}. \qquad (6.1)$$

Из формулы средней квадратической ошибки простой случайной выборки видно, что величина зависит от колеблемости признака в генеральной совокупности (чем больше вариация признака, тем больше ошибка выборки) и от объема выборки (чем больше обследуется единиц, тем меньше будет величина расхождений выборочных и генеральных характеристик).

Величину $t\mu_{\bar{x}}$ называют предельной ошибкой выборки. Обозначив предельную ошибку выборки как $\Delta_{\bar{x}}$, получим:

$$\Delta_{\bar{x}} = t\mu_{\bar{x}}, \qquad (6.2)$$

т.е. предельная ошибка выборки равна t-кратному числу средних ошибок выборки. Допустим, что $t = 2$. Тогда[9]

$$P\left(|\tilde{x} - \bar{x}| \le 2\mu_{\bar{x}}\right) = F_{t=2} = 0{,}9545,$$

т.е. с вероятностью, равной 0,9545, можно ожидать, что ошибка выборочной средней не превысит удвоенной средней квадратической ошибки выборки. Таким образом, величина предельной ошибки выборки может быть установлена с определенной вероятностью.

[9] Значения интегральной функции Лапласа для различных значений t приводятся в таблицах Приложения 2.

Приведем наиболее часто употребляемые уровни доверительной вероятности и соответствующие значения t для выборок достаточно большого объема ($n \geq 30$).

T	1,00	1,96	2,00	2,58	3,00
$\Phi(t)$	0,6827	0,9500	0,9545	0,9901	0,9973

Как видно из последней графы, вероятность появления ошибки, величина которой равна или больше утроенной средней ошибки выборки, т.е. $\Delta_{\bar{x}} \geq 3\mu_{\bar{x}}$, крайне мала и равна 0,0027 (1 − 0,9973). Такие маловероятные события считаются практически невозможными, а потому величину $\Delta_{\bar{x}} \geq 3\mu_{\bar{x}}$ можно принять за предел возможной ошибки выборки.

Ранее было доказано, что дисперсия доли σ^2_p равна $p(1-p)$, поэтому величину средней стандартной ошибки выборочной доли можно определить по следующей формуле:

$$\mu_p = \sqrt{\frac{p(1-p)}{n}}, \qquad (6.3)$$

а предельную ошибку доли по формуле:

$$\Delta_p \geq t\mu_p,$$

где p — генеральная доля (доля единиц, обладающих данным значением признака, в генеральной совокупности).

В формулы средней ошибки выборки (6.1) и (6.3) входит дисперсия признака в генеральной совокупности, величина которой, как правило, при проведении выборочного наблюдения неизвестна. Нам приходится использовать выборочную дисперсию в качестве оценки генеральной дисперсии.

Можно доказать, что $\sigma^2 = S^2 \frac{n}{n-1}$, где S^2 — выборочная дисперсия (дисперсия признака по результатам выборочного наблюдения).

Следовательно, генеральную дисперсию можно оценить по результатам одной выборки по следующей формуле:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1},$$

где $(n-1)$ — число степеней свободы для расчета дисперсии по выборочным данным[10].

Применение простой случайной выборки на практике весьма ограниченно. Прежде всего, практически нецелесообразно, а иногда невозможно повторное наблюдение одних и тех же единиц, а поэтому однажды обследованная единица повторному учету не подвергается. Применение бесповторного отбора взамен повторного диктуется также требованием повышения степени репрезентативности выборки (особенно при недостаточно больших n).

Средняя квадратическая ошибка случайной бесповторной выборки определяется по формулам[11]:

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)}; \qquad (6.4)$$

$$\mu_p = \sqrt{\frac{p(1-p)}{n}\left(1 - \frac{n}{N}\right)}. \qquad (6.5)$$

Сопоставление формул (6.1) и (6.4), (6.3) и (6.5) свидетельствует о том, что применение бесповторного отбора взамен повторного приводит к уменьшению стандартной ошибки выборки.

В тех случаях, когда численность генеральной совокупности N очень велика по сравнению с числом отобранных единиц n, величина $1 - (n/N)$ будет близка к единице, а потому ею можно пренебречь. Тогда ошибку случайного бесповторного отбора определяют по формуле простой случайной повторной выборки, что повышает надежность оценок генеральных характеристик по выборочным данным.

[10] При большом объеме выборки ($n > 30$) разница между дисперсиями, вычисленной по формуле $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$ и по формуле $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$, очень невелика, и поправка к выборочной дисперсии необязательна (так уже для $n = 30$ получим 30/(30−1) = 1,035, т.е. выборочная дисперсия будет на 3,5% меньше генеральной).

[11] Эти формулы являются упрощением строгой формулы $\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)}$, так как при достаточно больших N: $\frac{N-n}{N-1} = 1 - \frac{n}{N}$.

6.4. Как оценить генеральную среднюю и генеральную долю по данным выборки?

Выборочное наблюдение проводится в целях распространения выводов, полученных по данным выборки, на генеральную совокупность. Одной из основных задач является оценка по данным выборки интересующих нас характеристик (параметров) генеральной совокупности. Рассмотрим определение величины средней арифметической генеральной совокупности на основе выборочных данных.

Выборочное наблюдение дает возможность определить среднюю арифметическую выборочной совокупности $\bar{x}$ и величину предельной ошибки этой средней $\Delta_{\bar{x}}$, которая показывает (с определенной вероятностью), насколько выборочная средняя может отличаться от генеральной средней в большую или меньшую сторону. Тогда величина генеральной средней будет представлена интервальной оценкой, для которой нижняя граница будет равна $\bar{x} - \Delta_{\bar{x}}$, а верхняя граница — $\bar{x} + \Delta_{\bar{x}}$. Пределы, в которых с данной степенью вероятности будет заключена неизвестная величина оцениваемого параметра, называют *доверительными*, а вероятность P — *доверительной вероятностью*.

Доверительный интервал для генеральной средней можно записать так:

$$\bar{x} - t\mu_{\bar{x}} \le \bar{x} \le \bar{x} + t\mu_{\bar{x}}. \qquad (6.6)$$

Чаще всего доверительную вероятность устанавливают равной 0,95 или 0,99 (величины коэффициентов доверия t равны соответственно 1,96 и 2,58).

На рис. 6.1 показаны доверительные интервалы для четырех различных результатов выборочных наблюдений, построенных при вероятности 0,9545. Для каждой выборки под номерами 1, 2, 3 и 4 доверительный интервал для $\bar{x}$ будет определен так:

$$\bar{x}_j - 2\mu_{\bar{x}} \le \bar{x} \le \bar{x}_j - t\mu_{\bar{x}},$$

где $\bar{x}_j$ — средняя арифметическая j-й выборки.

Только доверительный интервал, построенный для выборочной средней, не включает генеральную среднюю.

Вероятность того, что величина генеральной средней выйдет за доверительные границы, будет равна $1 - P$, т.е. при $P = 0{,}9545$ она будет равна 0,0455 или 0,01 при $P = 0{,}99$. Событие, обладающее столь малой вероятностью, считается практически невозможным.

Аналогичным образом могут быть записаны доверительные пределы генеральной доли:

$$\omega - t\mu_{\bar{x}} \le \bar{x} \le \omega + t\mu_{\bar{x}}. \qquad (6.7)$$

Величина доверительного интервала для генеральной средней, или генеральной доли, зависит от величины предельной ошибки выборки $\Delta_{\bar{x}}$ или Δ_p. Чем больше величина предельной ошибки выборки, тем больше величина доверительного интервала, и тем, следовательно, ниже точность оценки.

Поскольку величина предельной ошибки равна $t\mu_{\bar{x}}$, точность оценки параметров генеральной совокупности будет зависеть от принятого уровня доверительной вероятности и от величины стандартной ошибки выборки. Допустим, что доля брака по данным выборки w составила 0,02 (или 2%) при $n = 1000$. Известно, что выборка равна 10% и, следовательно, $N = 10\,000$. Тогда величина стандартной ошибки бесповторной выборки равна 0,0042:

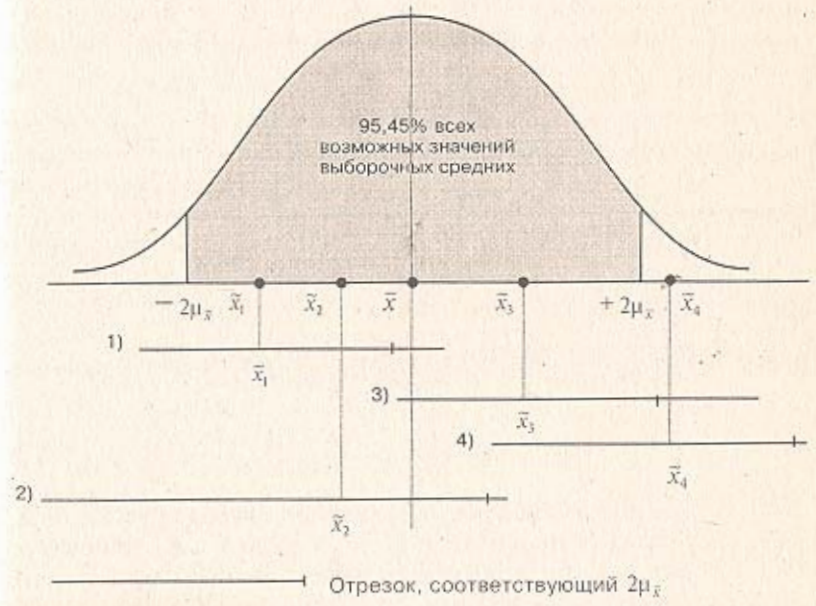


Рис. 6.1. Доверительные интервалы выборочных средних (P = 0,9545)

$$\mu_p = \sqrt{\frac{\omega(1-\omega)}{n}\left(1-\frac{n}{N}\right)} = \sqrt{\frac{0{,}02 \cdot 0{,}98}{1000}\left(1-\frac{1000}{10\,000}\right)} = 0{,}0042,$$

или 0,42%.

С вероятностью 0,954 мы можем утверждать, что предельная ошибка доли брака не превысит 0,0084 ($\Delta_p = 2\mu_{\bar{x}}$) и доля брака в генеральной совокупности будет находиться в интервале:

$0{,}0116 < p < 0{,}0284$, т.е. будет составлять от 1,16 до 2,84%.

Если мы примем доверительную вероятность равной 0,997, то величина предельной ошибки выборки составит 0,0126 (0,0042×3), и доля брака в генеральной совокупности будет находиться в интервале $0{,}02 \pm 0{,}0126$, т.е. $0{,}74\% \le p \le 3{,}26\%$.

Таким образом, с вероятностью 99,7% можно ожидать, что количество бракованных деталей во всей партии из 10 000 шт. будет находиться в интервале от 74 до 327 шт., тогда как с вероятностью 95,4% доверительный интервал составит от 116 до 284 шт. Изменение точности оценки генеральной доли в связи с изменением доверительной вероятности показано на рис. 6.2.

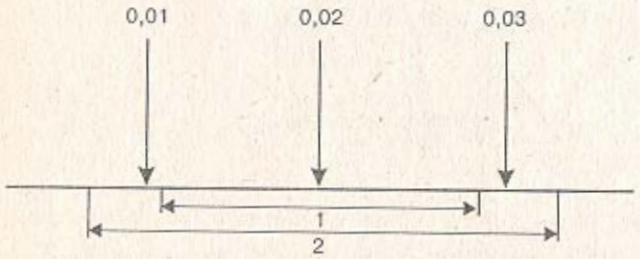


Рис. 6.2. Доверительные интервалы генеральной доли: (1) — вероятность P = 0,954; (2) — вероятность P = 0,997

Незначительно увеличивая достоверность выводов (с 95,4% до 99,7%), мы существенно снижаем точность оценки. В этой связи в экономических расчетах чаще рекомендуется использовать доверительную вероятность $P = 0{,}95$ или $P = 0{,}954$ (соответственно t равно 1,96 и 2,00).

6.5. Как определить необходимую численность выборки?

Средняя квадратическая (стандартная) ошибка выборки зависит от объема выборки и степени вариации признака в генеральной совокупности. Уменьшение стандартной ошибки выборки, а следовательно, увеличение точности оценки, всегда связано с увеличением объема выборки. В этой связи уже на стадии организации выборочного наблюдения приходится решать вопрос о том, каков должен быть объем выборочной совокупности, чтобы была обеспечена требуемая точность результатов наблюдений. Рассмотрим формулу предельной ошибки выборки для случая простой случайной повторной выборки:

$$\Delta_{\bar{x}} = t\mu_{\bar{x}} = t\sqrt{\frac{\sigma^2}{n}} = \frac{t\sigma}{\sqrt{n}}.$$

Отсюда

$$n = \frac{t^2\sigma^2}{\Delta_{\bar{x}}^2}. \qquad (6.8)$$

При проектировании выборочного наблюдения предполагаются заранее заданными величина допустимой ошибки выборки в соответствии с задачами конкретного исследования (следует иметь в виду, что в данной формуле $\Delta_{\bar{x}}$ — абсолютная величина предельной ошибки выборки) и вероятность выводов по результатам наблюдения (величина коэффициента t соответствует принятому уровню доверительной вероятности).

Величина σ^2, характеризующая дисперсию признака в генеральной совокупности, зачастую бывает неизвестна. Поэтому используют приближенные способы оценки генеральной дисперсии.

1. Можно провести «пробное» обследование (обычно небольшого объема), на базе которого определяется величина дисперсии признака, используемая в качестве оценки генеральной дисперсии:

$$\sigma^2 = \frac{\sum (x_i - \bar{x}_{проб})^2}{n_{проб} - 1},$$

где $\bar{x}_{проб}$ — средняя арифметическая по результатам пробного обследования;

$n_{проб}$ — число единиц, попавших в пробное обследование.

По данным нескольких пробных обследований выбирается наибольшее значение дисперсии.

2. Можно использовать данные прошлых выборочных обследований, проводившихся в аналогичных целях, т.е. дисперсия, полученная по их результатам, используется в качестве оценки генеральной дисперсии.

3. Если распределение признака в генеральной совокупности может быть отнесено к нормальному закону распределения, то размах вариации примерно равен 6σ (крайние значения отстоят в ту и другую сторону от средней на расстоянии 3σ), т.е. $R = 6\sigma$, откуда $\sigma \equiv 1/6R$, где $R = x_{max} - x_{min}$.

При проведении социально-экономических исследований, как правило, можно с достаточной точностью указать максимально и минимально возможные значения признака в исследуемой совокупности. Например, сколько операционистов нужно обследовать в банках региона, чтобы получить характеристику среднего уровня оплаты труда этой категории работников банков в регионе? Предложим, нам известно, что разница между наивысшим и наименьшим уровнем оплаты труда операциониста в регионе составляет 300 руб. Для нормального распределения в промежуток $x \pm 3\sigma$ включается 99,7% всех вариантов значений признака, а это означает применительно к рассматриваемой задаче, что 300 руб. примерно равно шести средним квадратическим отклонениям ($300 \approx \sigma$). Поэтому примерная оценка среднего квадратического отклонения заработной платы в генеральной совокупности операционистов региона составит 50 руб. ($\sigma = 300/6$). Для дальнейших расчетов достаточно, чтобы с вероятностью 0,954 предельная ошибка выборки не превышала 10 руб. Тогда, зная, что $\sigma = 50$ руб., а $t = 2$, и используя формулу (6.8) для определения необходимого объема выборки, получим:

$$n = \frac{2^2 50^2}{10^2} = 100 \text{ человек.}$$

Таким образом, при заданных условиях нужно обследовать размер заработной платы у 100 операционистов региона.

В связи с тем что генеральная дисперсия оценивается приближенно, рекомендуется величину объема выборки, рассчитанную по формуле (6.8), округлять в большую сторону. Эту же формулу рекомендуется использовать для определения объема выборки и в тех случаях, когда проводится случайная бесповторная выборка, так как всегда необходим некоторый «запас» числа обследованных единиц для обеспечения требуемой точности результатов.

Нередко на практике задается не величина абсолютной предельной ошибки $\Delta_{\bar{x}}$, а величина относительной погрешности $\Delta_{относ}$, выраженная в процентах к средней:

$$\Delta_{относ} = \frac{\Delta_{\bar{x}}}{\bar{x}} \cdot 100\%,$$

откуда

$$\Delta_{\bar{x}} = (\Delta_{относ} \cdot \bar{x})/100\%.$$

Подставив величину $\Delta_{\bar{x}}$, выраженную через относительную погрешность, в формулу (6.8), получим следующее выражение для определения необходимого объема выборки:

$$n = \frac{t^2\sigma^2}{\Delta^2_{относ}(\bar{x})^2} \cdot 100^2.$$

Как известно, отношение $\frac{\sigma}{\bar{x}} \cdot 100\%$ представляет собой коэффициент вариации V, откуда $n = \frac{t^2V^2}{\Delta^2_{относ}}$.

Например, по данным пробного обследования коэффициент вариации составляет 40%. Сколько нужно отобрать единиц, чтобы с вероятностью 0,954 предельная относительная ошибка выборки не превышала 5%?

При $V = 40\%$, $\Delta_{относ} = 5\%$ и $t = 2$

$$n = \frac{4 \cdot 1600}{25} = 256 \text{ ед.}$$

При определении по материалам выборки доли признака, а не средней его величины, объем выборочной совокупности определяется по следующим формулам:

а) для повторного отбора $n = \frac{t^2\omega(1-\omega)}{\Delta_p^2}$; (6.9)

б) для бесповторного отбора $n = \frac{t^2\omega(1-\omega)N}{N\Delta_p^2 + t^2\omega(1-\omega)}$. (6.10)

Для случая, когда частость w даже приблизительно неизвестна, можно произвести примерный расчет численности выборки, вводя в расчет максимальную величину дисперсии доли, равную 0,25:

а) для повторного отбора:

$$n = \frac{t^2}{4\Delta_p^2}; \qquad (6.9а)$$

б) для бесповторного отбора:

$$n = \frac{t^2 N}{4N\Delta_p + t^2}. \qquad (6.10а)$$

Очень часто цель выборочного наблюдения сводится к изучению различных признаков, характеризующих данную совокупность. Причем зачастую по степени колеблемости эти признаки могут сильно отличаться друг от друга. Кроме того, для различных признаков может требоваться разная степень точности. В таких случаях при определении необходимого объема выборки следует ориентироваться на тот признак, который при наибольшей колеблемости обладает наименьшей величиной допустимой ошибки.

Например, при проведении выборочного наблюдения требуется определить средний вес изделия и долю изделий первого сорта в совокупности объемом 5000 ед. Устанавливаем требуемую точность: для среднего веса — 1 г; для доли — 5%. По данным сплошного обследования изделий первой партии, состоящей из 100 ед., определили дисперсию веса, равную 36, и дисперсию доли, равную 0,05. Требуется определить необходимую численность случайной бесповторной выборки, обеспечивающей заданную точность наблюдения с вероятностью 0,954.

Исходные данные:

а) для определения n, обеспечивающего заданную точность оценки среднего веса изделия, примем:

$N = 5000$; $\sigma_x^2 = 36$; $\Delta_{\bar{x}} = 1$; $P = 0{,}954$; $t = 2$, тогда

$$n = \frac{t^2\sigma^2}{N\Delta_{\bar{x}}^2 + t^2\sigma_x^2} = \frac{4 \cdot 36 \cdot 5000}{5000 \cdot 1 + 4 \cdot 36} = 140 \text{ шт.};$$

б) для определения необходимого объема выборки, обеспечивающего требуемую точность в определении доли изделий первого сорта, мы располагаем такими данными:

$N = 5000$; $\Delta_p = 0{,}05$; $\sigma_p^2 = 0{,}05$; $P = 0{,}954$; $t = 2$, тогда

$$n = \frac{t^2\sigma_p^2 N}{N\Delta_p^2 + t^2\sigma_p^2} = \frac{4 \cdot 0{,}05 \cdot 5000}{5000 \cdot 0{,}0025 + 4 \cdot 0{,}05} = 80 \text{ шт.}$$

Очевидно, что число изделий, обеспечивающее требуемую точность результатов выборочного наблюдения для определения доли

изделий первого сорта, недостаточно, чтобы обеспечить требуемую точность при определении среднего веса изделий. Поэтому следует произвести отбор из генеральной совокупности не менее 140 изделий.

В целом формула предельной ошибки выборочной средней (доли) позволяет решать следующие три группы задач:

1) определять предел возможной ошибки средней (доли), т.е. величину возможных отклонений показателей генеральной совокупности от показателей выборочной совокупности;
2) определять необходимую численность выборки, обеспечивающую требуемую точность, при которой пределы возможной ошибки не превысят некоторой заранее заданной величины;
3) определять вероятность того, что в проведенной выборке ошибка будет иметь заданный предел.

Для решения всех перечисленных задач необходимо знать дисперсию признака в генеральной совокупности σ^2 и в случае бесповторной выборки ее объем N.

Чтобы оценить предельную ошибку выборки, нужно дополнительно задать определенную вероятность заключений, или, что то же, величину коэффициента доверия t, и знать объем выборки n.

При расчете объема выборки, необходимого для обеспечения заданной точности результатов наблюдения, нужно знать предельную ошибку выборки и величину коэффициента доверия t. Определяя численность выборки и ее точность, следует учитывать, что чем больше абсолютный объем выборки, тем менее ощутимо влияет на точность результата включение в выборку дополнительных десятков и даже сотен единиц и тем больших затрат требует дальнейшее повышение точности.

Покажем, как меняется необходимое число наблюдений при заданной величине доверительной вероятности и допустимой предельной ошибке при определении доли признака в исследуемой совокупности (табл. 6.1). Необходимое число наблюдений рассчита-

Таблица 6.1

Доверительная вероятность	Предельная ошибка (в %)			
	10	5	2	1
0,90	67	270	1690	6763
0,95	96	384	2400	9603
0,99	165	663	4146	16 587
0,997	220	880	5504	22 018

но исходя из формулы предельной ошибки повторного отбора при максимальной величине дисперсии доли, равной 0,25.

Из данных таблицы видно: чтобы гарантировать с вероятностью 0,95 величину отклонения от генеральной доли, не превышающую 10%, нужно обследовать 96 ед. Если предельную ошибку уменьшить до 5%, то нужно подвергнуть обследованию уже 384 ед., т.е. в четыре раза больше, а для обеспечения с вероятностью 0,95 точности в 1% при определении доли нужно обследовать уже 9603 ед. Таким образом, таблица позволяет исследователю сделать вывод о целесообразности увеличения объема выборки, а следовательно, о целесообразности дополнительных затрат, обеспечивающих то или иное повышение точности. При окончательном принятии решения об объеме выборки следует принимать во внимание не только статистическую точность получаемых результатов, но и учитывать финансовые ограничения.

Если задана предельная ошибка выборки D и объем выборки n, то можно определить величину коэффициента t, зная которую, по таблице Приложения 2, можно определить вероятность $F(t)$.

Вернемся к примеру, рассмотренному в параграфе 6.2. В выборке объемом 1000 ед. доля бракованных изделий составила 2%. Какова вероятность того, что во всей партии изделий (10 000 шт.) доля бракованных изделий будет находиться в пределах от 1,5 до 2,5%?

Доверительная вероятность, которую требуется определить, является функцией t, последняя находится из формулы предельной ошибки выборки $\Delta = t\mu_p$, откуда $t = \Delta_p / \mu_p$. Величина предельной ошибки выборки может быть определена как разность между максимально допустимой генеральной долей (по условию она равна 2,5%) и долей бракованных изделий в выборке (по условию 2%).

Таким образом, $\Delta_p = 0{,}5\%$ (2,5% − 2,0%). Так как выборка случайная бесповторная, то величина средней ошибки выборки находится по формуле:

$$\mu_p = \sqrt{\frac{\omega(1-\omega)}{n}\left(1-\frac{n}{N}\right)} = \sqrt{\frac{2\cdot 98}{1000}\left(1-\frac{1000}{10\,000}\right)} = 0{,}42\%.$$

Затем находим величину коэффициента доверия:

$$t = 0{,}5 / 0{,}42 = 1{,}19.$$

По таблицам интегральной функции Лапласа вероятность, соответствующая данной величине коэффициента t, равна 0,76595.

6.6. Каковы формы организации выборочного наблюдения?

Различные формы организации выборочного наблюдения — типическая (районированная или стратифицированная) выборка, серийная (гнездовая) выборка, механическая выборка, комбинированная выборка — представляют собой дальнейшее развитие и видоизменение простой случайной выборки. Применение этих форм обусловлено соображениями удешевления или облегчения процесса наблюдения, особым характером объекта наблюдения или отсутствием необходимой информации для составления списка единиц генеральной совокупности.

Рассмотрим организацию ***типической (стратифицированной) выборки***. Как правило, социально-экономические явления характеризуются большим разнообразием и не являются достаточно однородными в отношении изучаемых признаков. При наличии в составе генеральной совокупности различных типов явлений с разными уровнями изучаемых признаков желательно так организовать выборку, чтобы обеспечить более равномерное представительство в выборочной совокупности различных частей (типов) явления.

Для этого общий список единиц генеральной совокупности в целом предварительно разбивается на отдельные списки, каждый из которых включает единицы, принадлежащие к одной, однородной по определенному признаку группе (типу). В качестве типов могут быть использованы группы, сложившиеся в практике статистики. Например, при обследовании семейных бюджетов рабочих и служащих на первом этапе выделяют группы рабочих и служащих в отдельных отраслях экономики, затем при отборе работников на предприятии — квалифицированных и малоквалифицированных рабочих. Другими словами, типическая выборка опирается на статистическую группировку единиц генеральной совокупности по одному признаку или по комбинации нескольких признаков. Из каждой выделенной группы в случайном порядке отбирается некоторое количество единиц.

Таким образом, при проведении типической выборки необходимо разбить общий объем выборки n между группами и определить число подлежащих наблюдению единиц в каждой группе. Наиболее часто применяется так называемое ***пропорциональное распределение единиц***, когда количество отбираемых в выборку единиц пропорционально удельному весу данной группы по числу единиц в генеральной совокупности, т.е. число наблюдений по каждой группе определяется по формуле:

$$n_i = n\frac{N_i}{N}, \qquad (6.11)$$

где n_i — число наблюдений из i-й типической группы;

n — общий объем выборки;

N_i — объем i-й типической группы в генеральной совокупности;

N — объем генеральной совокупности.

Возможен и другой вариант (назовем его условно *равномерным распределением*), когда из каждой группы отбирают одинаковое число единиц:

$$n_i = \frac{n}{k}, \tag{6.11а}$$

где k — число выделенных типических групп.

Еще один вариант решения (назовем его условно *оптимальным размещением*) учитывает также степень вариации признака в различных группах генеральной совокупности. Расчет объема выборки из каждой типической группы производится по формуле:

$$n_i = n\frac{N_i\sigma_i}{\sum_{i=1}^{k} N_i\sigma_i}, \tag{6.11б}$$

где σ_i — среднее квадратическое отклонение изучаемого признака в i-й группе.

При таком размещении пропорция отбора для групп с большой колеблемостью признака увеличивается, что, в свою очередь, приводит к соответственному уменьшению возможной случайной ошибки при определении групповой средней. Использование принципа оптимального размещения на практике затрудняется тем, что часто мы не располагаем данными о величинах σ_i в генеральной совокупности.

При типическом отборе в выборку попадают таким образом представители всех типических групп, поэтому вероятность получить большую точность выборки здесь выше, чем при простой случайной выборке.

На основе типической выборки можно непосредственно определить средние по группам, т.е. величины $\bar{x}_i$. Величина средней для всей совокупности $\bar{x}_0$ есть средняя арифметическая взвешенная из выборочных средних по группам, причем в качестве веса используют объем соответствующей типической группы в генеральной совокупности N_i:

$$\bar{x}_0 = \frac{\sum_{i=1}^{k} \bar{x}_i N_i}{\sum_{i=1}^{k} N_i}. \tag{6.12}$$

При пропорциональном распределении единиц по типическим группам можно применять и формулу:

$$\bar{x}_0 = \frac{\sum_{i=1}^{k} \bar{x}_i n_i}{\sum_{i=1}^{k} n_i} = \frac{\sum_{i=1}^{k} \bar{x}_i N_i}{\sum_{i=1}^{k} N_i}. \tag{6.12а}$$

Из приведенной формулы очевидно, что точность типической выборки будет зависеть от степени надежности определения групповых средних, т.е. от того, насколько хорошо отобранные единицы представляют среднюю каждой типической группы, и следовательно, случайная ошибка типической выборки зависит от величины внутригрупповых дисперсий. Поэтому при определении ошибки выборки для средней величины признака в совокупности берется не общая дисперсия, а средняя из внутригрупповых дисперсий, которую обозначим $\bar{\sigma}^2$. Формулы для определения стандартной ошибки типической выборки для пропорционального и оптимального размещения приведены в табл. 6.14.

Так как общая дисперсия признака в генеральной совокупности согласно теореме сложения дисперсий равна сумме межгрупповой и средней из внутригрупповых дисперсий, величина стандартной ошибки типической выборки всегда меньше, чем при простой случайной выборке. Различия между общей дисперсией и средней из внутригрупповых будут тем больше, чем больше различия между группами (т.е. соответственно чем больше величина межгрупповой дисперсии). Таким образом, типическая выборка в сравнении с простой случайной выборкой имеет преимущества при наличии в составе генеральной совокупности резко различающихся между собой групп, в то время как колеблемость признака внутри групп незначительна.

Для организации типической выборки необходимо располагать данными, позволяющими выделить в генеральной совокупности однородные группы, составить списки единиц этих групп и т.д. Такие данные можно получить либо из материалов текущего оперативного учета, либо из данных сплошных переписей или учетов.

Таблица 6.4

Вид размещения единиц по типическим группам	Ошибка средней арифметической	Ошибка доли
	Способ отбора	Способ отбора
Пропорциональное	повторный $\sqrt{\frac{\sigma^2}{n}}$	повторный $\sqrt{\frac{\omega_i(1-\omega_i)}{n}}$
	бесповторный $\sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)}$	бесповторный $\sqrt{\frac{\omega_i(1-\omega_i)}{n}\left(1-\frac{n}{N}\right)}$
Оптимальное	повторный $\frac{1}{N}\sqrt{\sum\frac{\sigma_i^2 N_i^2}{n_i}}$	повторный $\frac{1}{N}\sqrt{\sum\frac{\omega_i(1-\omega_i)N_i^2}{n_i}}$
	бесповторный $\frac{1}{N}\sqrt{\sum\frac{\sigma_i^2 N_i^2}{n_i}\left(1-\frac{n}{N}\right)}$	бесповторный $\frac{1}{N}\sqrt{\sum\frac{\omega_i(1-\omega_i)N_i^2}{n_i}\left(1-\frac{n}{N}\right)}$

Примечание: σ_i^2 — дисперсия признака в соответствующей типической группе; ω_i — доля единиц, обладающих исследуемым признаком, в соответствующей типической группе.

При ***серийной выборке*** в случайном порядке отбираются группы (серии, гнезда) единиц, которые подвергаются сплошному обследованию. Серийная выборка перед уже рассмотренными видами выборки имеет преимущества организационного характера и широко используется там, где генеральная совокупность состоит из определенным образом обособленных групп единиц.

Так как внутри серии производится сплошное обследование всех входящих в нее единиц, средние по сериям определяются без случайной ошибки. Таким образом, общая средняя $\bar{x}$, рассматриваемая как оценка генеральной средней, будет зависеть от того, насколько точно серийные средние будут представлять генеральную среднюю. Тем самым случайная ошибка средней по всей выборке

зависит от величины межсерийной дисперсии σ^2, характеризующей степень колеблемости серийных средних[12].

При равновеликих сериях стандартная ошибка выборки определяется по формулам:

а) при повторном отборе серий:

$$\mu_{\bar{x}} = \sqrt{\frac{\delta^2}{s}}; \qquad (6.13)$$

б) при бесповторном отборе серий:

$$\mu_{\bar{x}} = \sqrt{\frac{\delta^2}{s}\left(\frac{S-s}{S-1}\right)} = \sqrt{\frac{\delta^2}{s}\left(1-\frac{s}{S}\right)}, \qquad (6.13а)$$

где S — общее число серий в генеральной совокупности;

s — число отобранных серий.

Из приведенных формул следует, что случайная ошибка серийной выборки должна быть тем меньше, чем меньше межсерийная дисперсия и чем, следовательно, более однородны серии внутри. Однако на практике крайне редко можно встретить такой вариант, когда отобранные в случайном порядке серии не отличаются одна от другой по величине изучаемого признака, тогда как единицы, входящие в отобранную серию, обычно близки по величине изучаемого признака. Это связано с общностью условий их формирования, что обусловливает малую величину внутригрупповой дисперсии.

Поскольку при серийной выборке, как правило, обследуется незначительное число серий, случайная ошибка серийной выборки получается несколько больше, чем при других способах отбора, т.е. серийная выборка менее точна, чем выборка, основанная на индивидуальном отборе. Тем не менее серийный отбор широко применяется в практике выборочных обследований, особенно в тех случаях, когда обследование охватывает обширную территорию и гнездами являются территориальные единицы, что связано со значительной экономией денежных средств на проведение наблюдения.

Средняя ошибка выборочной доли при случайной выборке в случае равновеликих серий определяется по формулам:

а) при повторном отборе:

[12] При равновеликих сериях $\bar{\bar{x}}_0 = \frac{\sum_{i=1}^{s}\bar{x}_i}{s}$, $\delta^2 = \frac{\sum_{i=1}^{s}(\bar{x}_i - \bar{\bar{x}}_0)^2}{s}$, где S — число отобранных серий.

$$\mu_p = \sqrt{\frac{\delta_p^2}{s}}, \qquad (6.14)$$

где δ^2 — межсерийная (межгрупповая) дисперсия доли, определяемая по формуле:

$$\delta_p^2 = \frac{\sum_{i=1}^{S} (\omega_i - \overline{\omega})^2}{S},$$

где ω_i — доли признака i-й серии;

$\overline{\omega}$ — доля признака по всей выборочной совокупности;

б) при бесповторном отборе:

$$\mu_p = \sqrt{\frac{\delta_p^2}{s}\left(1 - \frac{s}{S}\right)}. \qquad (6.14а)$$

6.7. В чем состоят преимущества механической выборки и как определяется величина ее ошибки?

Одной из распространенных форм в практике выборочного наблюдения является механический отбор. При механическом отборе наблюдению подвергаются единицы, находящиеся на равном расстоянии в определенной последовательности расположения единиц генеральной совокупности. Допустим, что имеется исчерпывающий список единиц генеральной совокупности и единицы располагаются в порядке, являющемся случайным по отношению к подлежащим изучению признакам (например, список рабочих по алфавиту). В зависимости от объема выборки из списка для обследования выбирается либо каждая четвертая единица, либо каждая десятая единица и т.д. При проведении механической выборки генеральная совокупность, в сущности, разбивается на равные по численности группы и из каждой группы для обследования отбирается одна единица.

В том случае, когда к механическому отбору прибегают в целях повышения репрезентативности выборочного наблюдения, списки единиц генеральной совокупности составляются в форме ранжированной совокупности, т.е. единицы заносятся в списки в порядке

возрастания (убывания) какого-либо признака, тесно связанного с изучаемыми признаками и до известной степени определяющего их величину. Например, при изучении бюджетов рабочих и служащих используется механический отбор из списков, составленных в порядке убывания величины средней месячной заработной платы.

Механический отбор очень удобен и в тех случаях, когда мы не можем заранее составить список единиц генеральной совокупности. Например, выборка берется из постепенно формирующейся во времени совокупности или из практически бесконечной совокупности. Так, при исследовании процесса покупок можно наблюдать каждого десятого входящего в магазин покупателя; при контроле качества продукции проверять, например, каждую пятую сходящую со станка деталь и т.д.

При проведении механической выборки нужно установить шаг отсчета, т.е. расстояние между отбираемыми единицами, и начало отсчета, т.е. номер той единицы, которая должна быть обследована первой. Шаг отсчета устанавливают в зависимости от предполагаемого процента отбора. Например, из генеральной совокупности объемом 1000 ед. обследованию подлежат 100 ед. Это означает, что из каждых 10 ед. обследованию будет подлежать одна единица, т.е. шаг отсчета равен 10.

Выбор начала отсчета непосредственно связан со способом расположения единиц генеральной совокупности в списках. В случае неупорядоченного расположения единиц из совокупности единиц первого интервала путем случайного отбора выбирают начальную единицу. Например, на основании жеребьевки номер начальной единицы равен 4. Тогда в выборку попадут единицы, стоящие в списке под номерами 4, 14, 24, 34,...., 984, 994.

Если единицы в списке ранжированы по определенному признаку, за начало отсчета принимают единицу, лежащую в середине первого интервала. В данном примере из первых десяти единиц нужно выбрать либо пятую, либо шестую единицу. Тогда в выборку попадут единицы с порядковыми номерами 5, 15, 25, 35, 45,...., 985, 995, (или 6, 16, 26, 36, 46,...., 986, 996).

Если генеральная совокупность предварительно не ранжируется и предполагается, что расположение единиц в ней является случайным по отношению к изучаемым признакам, то механический отбор можно рассматривать как разновидность простой случайной бесповторной выборки. Таким образом, величина случайной ошибки такого отбора будет определяться по формуле:

$$\mu_{\overline{x}} = \sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)}. \qquad (6.15)$$

В совокупностях бесконечно большого объема отношение $\frac{n}{N}$ очень мало и множитель $\left(1-\frac{n}{N}\right)$ очень мало отличается от единицы. Поэтому для практически бесконечных совокупностей величина случайной ошибки механической выборки может быть определена по формуле:

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}}. \qquad (6.15а)$$

При механическом отборе из ранжированной совокупности единицы отбора расположены не в случайном порядке, а в порядке, обеспечивающем более пропорциональное представительство в выборке отдельных групп генеральной совокупности, различающихся по величине признака. Такой отбор можно уподобить типической выборке с пропорциональным размещением. Генеральная совокупность при этом способе отбора как бы разбивается на однородные группы равного объема и из каждой группы выбирается одна единица. Однако такое сопоставление механического отбора из ранжированной совокупности с типической выборкой не вполне правомерно, поскольку границы групп устанавливаются в известной мере произвольно, механически, кроме того, определение ошибки по формулам типической выборки практически невозможно из-за отсутствия данных о вариации по группам. В этой связи величина ошибки механического отбора из ранжированной совокупности приближенно оценивается по формулам простой случайной выборки.

Практика организации выборочных наблюдений часто опирается на многоступенчатый отбор. Многоступенчатый отбор характеризуется тем, что на всех ступенях, кроме последней, осуществляется только отбор, а наблюдение единиц производится только на последней ступени. Поэтому списки единиц нужны лишь на последней ступени отбора. При многоступенчатой выборке единицы отбора на первых ступенях обычно представляют собой организационные единства единиц наблюдения и на разных ступенях применяются единицы отбора разных порядков.

В качестве примера рассмотрим принципиальную схему построения репрезентативно территориальной выборки, реализуемой Всероссийским центром изучения общественного мнения (ВЦИОМ). Такая выборка, используемая в проекте «Всероссийский мониторинг в социальной сфере», представляет собой многоступенчатую

территориальную стратифицированную выборку. Предварительная стратификация включала следующие этапы (ступени).

На первой ступени стратификации вся территория России была разделена на экономико-географические регионы, где в качестве дополнительного фактора учитывалась национальная структура населения (наличие или отсутствие национальных автономий).

На втором этапе в качестве стратообразующих признаков использовались численность населения и административный статус единиц отбора. С учетом того, что было выделено 18 регионов и 12 групп на втором этапе, максимально возможное число страт составила 216 (18×12).

Фактическое же число страт было значительно ниже, так как были исключены из-за разных обстоятельств некоторые регионы (например, в силу чрезвычайных обстоятельств некоторые регионы Северного Кавказа), а также некоторые страты, выделенные на втором этапе, отсутствовали в ряде регионов. (Например, областные центры с численностью населения более 1 млн человек и т.д.)

Следующим шагом на стадии проектирования выборочного наблюдения являлось распределение общего объема выборки между стратами. Для этого подсчитывалось число постоянно проживающего взрослого населения каждой страты и объем выборки в каждой из них являлся пропорциональным доле взрослого населения в этой страте.

Окончательный этап формирования выборочной совокупности учитывал и ограничение на минимальное число опрашиваемых в каждой страте. Так, в 1997 г. минимальное число было равно 10. Поэтому при объеме выборки 2100 ед. приходилось объединять те страты, где население составляло менее 0,5% взрослого населения Российской Федерации $\left(\frac{10}{2100} \cdot 100\%\right)$.

С учетом всех вышеназванных обстоятельств окончательное количество страт в обследовании 1997 г. составило 66.

В составе каждой страты производился случайный отбор городских населенных пунктов и сельских административных районов. И наконец, на последней ступени отбора перед формированием списка домохозяйств отбирались избирательные участки (в городских населенных пунктах) и деревни (в сельских районах).

Затем составлялись списки всех взрослых жителей домохозяйств в отобранных на последней ступени единицах отбора. Из этих списков производился отбор респондентов.

В целях экономии средств данные по некоторым интересующим нас признакам можно анализировать на основании изучения всех

единиц выборочной совокупности, а по другим признакам — на основании части единиц выборочной совокупности, которые представляют подвыборку из единиц первоначальной выборки. Этот способ называют *двухфазным отбором*. При наличии нескольких подвыборок можно говорить о *многофазном отборе*.

Многофазный отбор по своей структуре отличается от многоступенчатого отбора, так как при многофазном отборе пользуются на каждой фазе одними и теми же единицами отбора, при многоступенчатом отборе на разных ступенях применяются единицы отбора разных порядков. Многофазным отбором чаще всего пользуются в тех случаях, когда различно число единиц, необходимых для определения отдельных показателей с заданной точностью. Это связано как с различиями в степени колеблемости взаимосвязанных переменных, так и с разной точностью, требуемой для подсчетов. Ошибки при многофазной выборке рассчитывают на каждой фазе отдельно.

Иногда бывает целесообразно взять из изучаемой совокупности две или несколько независящих друг от друга выборок, применяя для получения каждой из них один и тот же способ отбора. Такие выборки называют ***взаимопроникающими выборками***. Преимущество взаимопроникающих выборок состоит в том, что они позволяют получить отдельные и независимые оценки тех или иных признаков изучаемой совокупности. Если проведение разных выборок поручено разным исследователям, то сравнение результатов обеспечивает взаимную проверку работы самих исследователей.

6.8. Что такое малые выборки?

Как уже указывалось, для определенного способа отбора единиц величина стандартной ошибки зависит от объема выборки и степени колеблемости изучаемого признака в генеральной совокупности. Причем, чем меньше объем выборки, тем большую величину стандартной ошибки следует ожидать, а это, в свою очередь, снижает точность оценки параметров генеральной совокупности.

При выборках небольшого объема величина выборочной дисперсии, используемой в качестве оценки дисперсии генеральной совокупности, может быть в значительной мере подвержена влиянию случайностей выборки. Поэтому при выборках небольшого объема методы оценки результатов выборочного наблюдения видоизменяются в сравнении с применяемыми в теории большими выборками. (К безусловно малым относятся выборки объемом менее 30 ед.)

Первые работы в области теории малой выборки были выполнены английским статистиком В. Госсетом в 1908 г. (его псевдоним — Стьюдент) и продолжены в исследованиях Р. Фишера.

Для оценки возможных пределов ошибки малой выборки пользуются так называемым отношением Стьюдента:

$$t = \frac{\tilde{x} - \bar{x}}{S : \sqrt{n-1}}. \qquad (6.16)$$

Знаменатель приведенной формулы представляет собой меру случайных колебаний выборочной средней в малой выборке $\mu_{м.в.}$:

$$\mu_{м.в.} \frac{S}{\sqrt{n-1}}, \qquad (6.16а)$$

где S — величина среднего квадратического отклонения для данной выборки ($S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \tilde{x})^2}{n}}$ не рассматривается как оценка генеральной дисперсии).

Таким образом, теоретическое распределение отношения Стьюдента t имеет дело с величинами, определяемыми непосредственно по данным выборки.

Закон распределения величины t носит название «закон распределения Стьюдента», согласно которому плотность распределения ошибок малой выборки $Sk(t)$ равна:

$$Sk(t) = A\left(1 + \frac{t^2}{k}\right)^{\frac{k+1}{2}},$$

где k — число степеней свободы варьирования при определении выборочной дисперсии, равное $(n-1)$;

A — определяется в зависимости от числа степеней свободы с помощью гамма-функции (Г-функции):

$$A = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{\pi k}}.$$

Вероятность того, что ошибка выборки будет не больше заданной величины $\mu_{м.в.}$ определяется интегральной функцией:

$$Sk(t) = A\int_{-\infty}^{t}\left(1+\frac{t^2}{k}\right)^{-\frac{k+1}{2}} dt.$$

Следует учитывать, что закон Стьюдента приложим только к оценке ошибок выборок, взятых из генеральной совокупности с нормальным распределением признака.

Кривая t-распределения симметрична относительно оси ординат. Как видно из приводимой функции плотности, координаты кривой распределения Стьюдента зависят не только от t, но и от числа степеней свободы k.

Для больших значений k (больше 30) кривая $Sk(t)$ очень близка к кривой нормального распределения. Но при малом числе степеней свободы она значительно отклоняется от кривой нормального распределения, медленнее спускаясь к оси абсцисс, т.е. значительные отклонения от средней имеют в данном случае большую вероятность, чем для нормального распределения. Следовательно, вероятность больших ошибок в малой выборке выше, чем при больших выборках.

Составленные таблицы распределения Стьюдента облегчают применение отношения t к практике. Приведем выдержку из таблицы $Sk(t)$ для некоторых значений k и t (табл. 6.5).

Приведенная таблица показывает вероятность того, что фактическое отношение Стьюдента $t_ф$ в условиях случайной выборки будет не больше приведенного в таблице, т.е. $P(t_ф < |t|) = Sk(t)$. Но по этим же таблицам можно определить и вероятность того, что фактическое отношение не превзойдет по абсолютной величине табличное значение, т.е. $P(t_ф < |t|)$. Эта вероятность определяется по формуле: $P(t_ф < |t|) = Sk(t) - [1 - Sk(t)] = 2Sk(t) - 1$.

Например, при $k = 20$, вероятность того, что предельная ошибка выборки будет не более 3μ, равна 0,996 (см. табл. 6.5). Вероятность того, что ошибка будет заключена в пределах $\pm 3\mu$ будет равна: $2 \times 0{,}996 - 1 = 0{,}992$. Таблицу Стьюдента часто излагают в краткой удобной для практического использования форме — для каждого числа степеней свободы k указывают величину отношения Стьюдента t, которая с заданной вероятностью (0,95 или 0,99) не будет превышена по абсолютной величине в силу случайностей отбора (см. Приложение 4).

Таблица 6.5

t \ k	4	5	9	10	15	20	25	Нормальное распределение
1,0	0,813	0,818	0,828	0,830	0,833	0,835	0,838	0,841
2,0	0,942	0,949	0,962	0,963	0,968	0,970	0,973	0,977
3,0	0,980	0,985	0,992	0,993	0,995	0,996	0,997	0,999

Приведем пример, иллюстрирующий практическое применение распределения Стьюдента. Оздоровительный центр, рекламируя свои услуги, предлагает клиентам за короткий срок снижение веса до 10 кг. По результатам выборочного обследования 15 женщин, воспользовавшихся услугами центра, были получены следующие данные о снижении их веса:

№ п/п	Снижение веса (в кг)	№ п/п	Снижение веса (в кг)
1	10,2	9	5,2
2	7,6	10	6,1
3	6,1	11	5,0
4	8,4	12	3,7
5	6,0	13	4,7
6	5,7	14	3,6
7	13,7	15	3,2
8	6,9		

Выборочная средняя составляет 6,41 кг ($\tilde{x} = \frac{96{,}1}{15} = 6{,}41$ кг), т.е. среднее снижение веса у обследованных женщин составило 6,41 кг.

Выборочная дисперсия равна 7,061 $\left(S^2 = \frac{\sum_{i=1}^{n}(x_i - 6{,}41)}{15} = 7{,}061\right)$.

Следовательно, средняя квадратическая ошибка выборки составит 0,71 кг:

$$\mu_{м.в.}\sqrt{\frac{7{,}061}{14}} = 0{,}71 \text{ кг}.$$

Оценим с вероятностью 0,99 предел возможных расхождений выборочной средней и генеральной средней. Так как число степеней свободы равно 14 ($k = n - 1 = 15 - 1$), то по таблице Приложения 4 находим, что значение t, соответствующее вероятности 0,99, равно 2,977.

Тогда с вероятностью 0,99 можно предполагать, что ошибка выборочной средней будет не больше 2,114 кг (2,977 × 0,71), а снижение веса пациентов оздоровительного центра будет находиться в пределах от 4,3 до 8,52 кг (6,41 ±2,11). Следовательно, указанное в рекламе снижение веса на 10 кг имеет столь малую вероятность, что считается событием практически невозможным.

6.9. Каковы общие понятия о статистической проверке гипотез? Как выбирают тип критической области?

Под статистической гипотезой понимаются различного рода предположения относительно характера или параметров распределения случайной переменной, которые можно проверить, опираясь на результаты наблюдений в случайной выборке.

Следует иметь в виду, что статистическая проверка гипотез имеет вероятностный характер, так как принимаемые выводы основываются на изучении свойств распределения случайной переменной по данным выборки, а потому всегда существует риск допустить ошибку. Однако с помощью статистической проверки гипотез можно определить вероятность принятия ложного решения. Если вероятность последнего невелика, то можно считать, что применяемый критерий обеспечивает малый риск ошибки.

При проверке гипотез имеется возможность совершить ошибку двоякого рода:

а) ошибка первого рода — проверяемая гипотеза (ее обычно называют нулевой гипотезой и обозначают H_0) является в действительности верной, но результаты проверки приводят к отказу от нее;

б) ошибка второго рода — проверяемая гипотеза в действительности является ошибочной, но результаты проверки приводят к ее принятию.

Для построения статистического критерия, позволяющего проверить некоторую гипотезу, необходимо следующее:

1) сформулировать проверяемую гипотезу H_0. Наряду с проверяемой гипотезой формулируется также конкурирующая (альтернативная) гипотеза;

2) выбрать уровень значимости α, контролирующий допустимую вероятность ошибки первого рода;

3) определить область допустимых значений и так называемую критическую область;

4) принять то или иное решение на основе сравнения фактического и критического значений критерия.

Уровнем значимости будем называть такое малое значение вероятности попадания критерия в критическую область при условии справедливости гипотезы, что появление этого события может расцениваться как следствие существенного расхождения выдвинутой гипотезы и результатов выборки. Обычно уровень значимости принимают равным 0,05 или 0,01. Исходя из величины уровня значимости, можно определить ***критическую область***, под которой понимается такая область значений выборочной характеристики, попадая в которую они будут свидетельствовать о том, что проверяемая гипотеза должна быть отвергнута. К критической области относятся те значения, появление которых при условии верности гипотезы было бы маловероятным.

Допустим, что рассчитанное по эмпирическим данным значение критерия попало в критическую область, тогда при условии верности проверяемой гипотезы H_0 вероятность этого события будет не больше уровня значимости α. Поскольку уровень значимости α выбирается достаточно малым, то такое событие является маловероятным и, следовательно, проверяемая гипотеза H_0 может быть отвергнута.

Если же наблюдаемое значение характеристики не принадлежит к критической области и, следовательно, находится в области допустимых значений, то проверяемая гипотеза H_0 не отвергается. Вероятность попадания критерия в область допустимых значений при справедливости проверяемой гипотезы равна $1-\alpha$.

Чем меньше уровень значимости, тем меньше вероятность забраковать проверяемую гипотезу, когда она верна, т.е. меньше вероятность совершить ошибку первого рода. Но при этом расширяется область допустимых значений и, значит, увеличивается вероятность совершения ошибки второго рода.

При данном уровне значимости мы можем по-разному устанавливать критическую область, гарантирующую этот уровень. Пусть мы задаем уровень значимости $\alpha = 0{,}05$. В качестве критической области, отвечающей этому уровню значимости, мы можем взять:

1. Область больших положительных отклонений, при которых вероятность того, что значение x выйдет за пределы $\bar{x} + t\mu_{\bar{x}}$, равна уровню значимости α. При $\alpha = 0{,}05$ сказанное можно записать так:

$$P_1 = P\left(x > \bar{x} + t_1\mu_{\bar{x}}\right) = 0{,}05.$$

Тогда $P_1 = \frac{1}{\sqrt{2\pi}}\int_{t_1}^{\infty} e^{-\frac{t^2}{2}}dt = \frac{1}{2} - \frac{1}{\sqrt{2\pi}}\int_{0}^{t} e^{-\frac{t^2}{2}}dt$,

откуда $\frac{1}{\sqrt{2\pi}}\int_{0}^{t} e^{-\frac{t^2}{2}}dt = \frac{1}{2} - 0{,}05 = 0{,}45.$

По таблице нормированной функции Лапласа находим, что t_1, соответствующее вероятности 0,45, равно 1,64 (см. Приложение 2).

Теперь граница критической области будет найдена из выражения $x > \bar{x} + 1{,}64\mu_{\bar{x}}$. В данном случае мы имеем одностороннюю критическую область, графическое представление которой приведено на рис. 6.3.

2. Область больших отрицательных отклонений:

$P_2 = P\left(x < \bar{x} < t_2\mu_{\bar{x}}\right) = 0{,}05$, т.е. $P_2 = P\left(x < \bar{x} - 1{,}64\mu_{\bar{x}}\right)$.

3. Область больших по абсолютной величине отклонений (рис. 6.4):

$$P_3 = P\left(\left|x - \bar{x}\right| > t_3\mu_{\bar{x}}\right) = 0{,}05,$$



Рис. 6.3. Правосторонняя критическая область (область больших положительных отклонений)



Рис. 6.4. Двусторонняя критическая область при $\alpha = 0{,}05$

т.е. $0{,}05 = 1 - 2\cdot\frac{1}{\sqrt{2\pi}}\int_{0}^{t} e^{-\frac{t^2}{2}}dt$.

Тогда $\frac{1}{\sqrt{2\pi}}\int_{0}^{t} e^{-\frac{t^2}{2}}dt = \frac{1 - 0{,}05}{2} = 0{,}475.$

По таблице нормированной функции Лапласа находим, что $t_3 = 1{,}96$. Таким образом, $\left[P\left|x - \bar{x}\right| > 1{,}96\mu_{\bar{x}}\right] = 0{,}05$.

4. Область малых по абсолютной величине отклонений:

$$P_4 = P\left(\left|x - \bar{x}\right| < t_4\mu_{\bar{x}}\right) = 2\cdot\frac{1}{\sqrt{2\pi}}\int_{0}^{t} e^{-\frac{t^2}{2}}dt = 0{,}05.$$

Тогда $\frac{1}{\sqrt{2\pi}}\int_{0}^{t} e^{-\frac{t^2}{2}}dt = 0{,}025.$

Величина t, соответствующая вероятности 0,025, равна 0,063 (см. Приложение 2).

6.10. Как проверяется гипотеза о принадлежности «выделяющихся» наблюдений исследуемой генеральной совокупности?

В составе собранных данных могут встречаться единичные наблюдения, у которых зарегистрированные значения признака заметно отличаются от общего уровня. Происхождение таких отличий может быть различным. Они могут возникать в результате: ошибок наблюдения; случайного стечения различного рода обстоятельств, в отдельности несущественных; нарушения однородности изучаемой совокупности.

Субъективное отбрасывание вариантов ряда, даже резко отличающихся от средней арифметической, не имеет никакого принципиального оправдания, поскольку первичные данные соответствуют конкретным случаям массового явления. Для исключения из дальнейшей обработки выделяющихся наблюдений необходимо применение надлежащим образом обоснованных критериев.

Допустим, что распределение результатов наблюдений, проводимых в обычных условиях, следует нормальному закону с параметрами $\bar{x}$ и σ. При проведении одной из серий таких наблюдений мы получим n значений $x_1 \le x_2 \le \ldots \le x_n$, среди которых максимальное значение x_n (или минимальное x_1) резко отличается о своей величине от остальных $(n-1)$ наблюдений. Возникает вопрос, относятся ли эти значения к данной наблюдаемой в определенных условиях генеральной совокупности или они возникли как результат каких-то экстраординарных обстоятельств. Нулевой гипотезой в этом случае является предположение о том, что $x_1(x_n)$ принадлежит той же совокупности, что и все остальные $n-1$ наблюдений, т.е. x_n (или x_1) не является результатом ошибки наблюдения или изменения общих условий формирования уровня признаков совокупности.

Проверка этой гипотезы состоит в том, что $x_n(x_1)$ сравнивается по величине с некоторой критической границей x. Если выделяющимся наблюдением является наибольшее, то x_n сравнивается с верхней допустимой границей, выбранной таким образом, чтобы вероятность превзойти ее была равна уровню значимости, т.е. в данном случае мы имеем дело с критической областью вида:

$$P\left(x_n < \bar{x} + t\sigma\right) = \alpha .$$

Гипотеза H_0 бракуется, если x_n превосходит по величине указанную границу (см. рис. 6.4).

Если же выделяющимся наблюдением является наименьшее (x_1), то его сравнивают с нижней допустимой границей, которую принимают равной $\bar{x} - t\sigma$, т.е. $P\left(x_n < \bar{x} - t\sigma\right) = \alpha$.

Если же испытанию одновременно подлежат и максимальное и минимальное значения признака, то критическая область будет иметь вид: $P\left(\left|x - \bar{x}\right| > t\sigma\right) = \alpha$ (см. рис. 6.5).

Например, известны следующие данные, приведенные в табл. 6.7.

Используем критическую область вида $P\left(x_n > \bar{x} + t\sigma\right)$, так как минимальное значение x_1 мало отличается от следующего за ним значения в ранжированном ряду, т.е. проверим, принадлежит ли выделяющееся наблюдение $x_n = 178$ к рассматриваемой совокупности результатов.

При уровне значимости $\alpha = 0{,}01$ значение нормированной функции Лапласа для рассматриваемой критической области будет равно:

$$\frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{t^2}{2}} dt = \frac{1}{2} - 0{,}01 = 0{,}49.$$

Этому значению в таблице нормированной функции Лапласа соответствует $t = 2{,}33$. Тогда верхняя допустимая граница значений признака, которая не может быть превышена с вероятностью 0,99, будет равна 139,6 (57,3+2,33x35,3 = 139,6). Значение $x_n = 178$ выходит за рассчитанную границу, поэтому с вероятностью 0,99 можно считать $x_n = 178$ не принадлежащим к изучаемой совокупности, и следует исключить его из дальнейших расчетов.

Часто могут иметь место такие случаи, когда параметры генеральной совокупности $\bar{x}$ и σ неизвестны, и для проверки гипотезы о выделяющихся наблюдениях используются полученные по выборке оценки соответствующих параметров. Однако следует учиты-

Таблица 6.7

Количество наблюдений	Минимальные значения		Максимальные значения		Разность смежных значений		Среднее значение	Среднее квадратическое отклонение по выборке, σ
	x_1	x_2	x_{n-1}	x_n	$x_2 - x_1$	$x_n - x_{n-1}$		
84	7	13	130	178	6	48	57,3	35,3

вать, особенно для малых выборок, что эти оценки являются не вполне надежными. В этой связи для отбрасывания выделяющихся наблюдений по данным малых выборок можно рекомендовать пользоваться критерием Ф. Граббса, предложенным в 1950 г.

Критерий Ф. Граббса основан на отношении двух сумм квадратов отклонений:

1. Для испытания наибольшего наблюдения, являющегося выделяющимся в выборке объема n из нормально распределенной совокупности, рассчитывается отношение:

$$\frac{S_n^2}{S^2} = \frac{\sum_{i=1}^{n-1}(x_i - \bar{x}_n)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \le K, \qquad (6.17)$$

где $x_1 \le x_2 \le x_3 \le \ldots \le x_n$; $\bar{x}_n = \frac{\sum_{i=1}^{n-1} x_i}{n-1}$; $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$.

2. Для испытания наименьшего наблюдения, выделяющегося в выборке объема n, рассчитывается отношение:

$$\frac{S_1^2}{S^2} = \frac{\sum_{i=2}^{n}(x_i - \bar{x}_1)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \le K, \qquad (6.17а)$$

где $\bar{x}_1 = \frac{\sum_{i=2}^{n} x_i}{n-1}$.

Рассчитанная величина отношения ($K_{расч}$) сравнивается с табличной величиной ($K_{табл}$) при определенном числе наблюдений и заданном уровне значимости. $K_{табл}$ характеризует ту предельную величину расхождений в суммах квадратов отклонений, которая с вероятностью $(1 - \alpha)$ может быть объяснена случайными причинами.

Если $K_{расч}$ будет равно или меньше табличного, то наименьшее или наибольшее наблюдения не отбрасываются. Если же $K_{расч} > K_{табл}$, то вероятность того, что расхождения в суммах квадратов отклонений можно объяснить случайными причинами, равна уровню значимости α и в силу малой вероятности считается событием практически невозможным. В таком случае выделяющееся наблюдение следует отбросить и для дальнейших расчетов использовать оставшиеся $(n - 1)$ наблюдений.

Покажем применение критерия Ф. Граббса на следующем примере. Отклонения деталей от номинального размера оказались такими (в мм): 0,07; 0,09; 0,10; 0,12; 0,13; 0,15; 0,16; 0,17; 0,25.

Определим, не содержат ли результаты ошибки наблюдения, предполагая, что распределение размеров в генеральной совокупности следует нормальному закону.

Средняя арифметическая по всем отклонениям равна 0,138:

$$\bar{x} = \frac{0{,}07 + 0{,}09 + 0{,}10 + 0{,}12 + 0{,}13 + 0{,}15 + 0{,}16 + 0{,}17 + 0{,}25}{9}.$$

Исключая максимальное отклонение 0,25, получим:

$$\bar{x}_n = 0{,}124\left(\frac{0{,}07 + 0{,}09 + 0{,}10 + 0{,}12 + 0{,}13 + 0{,}15 + 0{,}16 + 0{,}17 + 0{,}25}{8}\right).$$

Сумма квадратов отклонений от $\bar{x}$ будет равна 0,0231:
$[(0{,}07 - 0{,}138)^2 + (0{,}09 - 0{,}138)^2 + \ldots + (0{,}17 - 0{,}138)^2 + (0{,}25 - 0{,}138)^2]$.

Сумма квадратов отклонений $(n - 1)$ значений от $\bar{x}$ равна 0,0088:
$[(0{,}07 - 0{,}124)^2 + (0{,}09 - 0{,}124)^2 + \ldots + (0{,}16 - 0{,}124)^2]$.

Таблица 6.8

Таблица Ф. Граббса (выдержка)

Число наблюдений n	Уровень значимости (α)	
	0,01	0,05
3	0,0001	0,0027
4	0,0100	0,0494
5	0,0442	0,1270
6	0,0928	0,2032
7	0,1447	0,2696
8	0,1948	0,3261
9	0,2411	0,3742
10	0,2831	0,4154
...		
15	0,4401	0,5559
...		
20	0,5393	0,6379
...		
25	0,6071	0,6923

Отношение двух сумм квадратов отклонений:

$K_{расч} = 0{,}0088/0{,}0231 = 0{,}381.$

При числе наблюдений $n = 9$ и уровне значимости 0,01 $K_{табл}$ равно 0,2411, т.е. $K_{расч} > K_{табл}$ (табл. 6.8). Полученное расчетное значение больше табличного и при уровне значимости 0,05 ($K_{табл} = 0{,}3742$), а поэтому отклонение от номинального размера 0,25 мм следует отнести к ошибкам наблюдения.

Рассмотренные выше критерии для отбрасывания выделяющихся наблюдений базировались на предположении нормального распределения признака в генеральной совокупности. На такого же рода допущениях о нормальности распределения признака основана и оценка существенности коэффициента корреляции и т.д.

Поэтому одним из этапов статистической обработки результатов наблюдений является проверка гипотезы о возможности отнесения эмпирического распределения, полученного по данным выборки, к типу кривых нормального распределения (см. гл. 5).

Еще один вариант проверки гипотезы о принадлежности выделяющихся наблюдений к генеральной совокупности исследуемых данных был предложен Дж. Ирвином в 1925 г. Критерий Ирвина λ основан на соотношении разности между x_n и x_{n-1} к среднему квадратическому отклонению в генеральной совокупности:

$$\lambda = \frac{x_n - x_{n-1}}{\sigma}, \qquad (6.18)$$

где $x_1 \le x_2 \le x_3 \le \ldots \le x_{n-1} \le x_n$; $\sigma = \sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$.

В нижеприводимой выдержке из таблицы Дж. Ирвина (табл. 6.9) приводится вероятность того, что разность между x_n и x_{n-1} отличается от среднего квадратического отклонения в случайной выборке объемом n единиц более, чем в λ раз.

Используя данные примера, приведенного в табл. 6.7, рассчитаем критерий Ирвина λ, который составит 1,36 (λ = 48 / 35,3 = 1,36). По данным табл. 6.9, при $n = 80$ и λ = 1,4 $P(\lambda)$ равно 0,015. В силу малой вероятности появления таких расхождений между смежными значениями в ранжированном ряду полученных данных можно считать значение $x_n = 178$ не принадлежащим к изучаемой совокупности и при дальнейшей обработке информации использовать оставшиеся 83 значения.

Таблица 6.9

Значения P(λ), где λ — критерий Ирвина

Объем выборки n	Значения вероятности при критерии λ, равном										
	1,0	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0
10	0,152	0,121	0,096	0,075	0,059	0,045	0,038	0,025	0,020	0,015	0,011
20	0,107	0,082	0,062	0,047	0,035	0,026	0,019	0,014	0,010	0,007	0,005
30	0,089	0,068	0,050	0,037	0,027	0,020	0,014	0,010	0,007	0,005	0,004
. . .											
60	0,065	0,048	0,034	0,025	0,017	0,012	0,009	0,006	0,004	0,003	0,002
70	0,061	0,044	0,032	0,022	0,016	0,011	0,008	0,005	0,004	0,002	0,002
80	0,058	0,041	0,030	0,021	0,015	0,010	0,007	0,005	0,003	0,002	0,001

6.11. Как проверяется гипотеза о величине средней арифметической и доли?

Эта гипотеза чаще всего проверяется при неизвестной генеральной дисперсии. Например, средний расход сырья на единицу продукции при существующем методе производства составляет 2,8 усл. ед. После внесения изменений в технологию изготовления продукции по результатам проверки достаточно большой партии изделий средний расход сырья на единицу продукции составил 2,6 усл. ед. Средняя ошибка выборки оказалась равной 0,1. Возникает вопрос, действительно ли применение нового метода обработки приводит к снижению материалоемкости продукции?

Нулевая гипотеза состоит в том, что между новым и существующим методами производства изделий отсутствуют различия, существенные с точки зрения их влияния на материалоемкость, т.е. между генеральными средними при старом и новом методах производства нет значительной разницы, а отклонение выборочной средней от достигнутого уровня при действующем методе обусловлено только случайностями выборки, т.е. H_0 означает, что $\bar{x}_c = \bar{x}_н$, где $\bar{x}_c$ и $\bar{x}_н$ — средний расход сырья на единицу продукции соответственно при существующем и новом методах производства.

Альтернативная гипотеза может быть сформулирована двояко:

1. Применение нового метода обработки приводит к изменению расхода сырья на единицу продукции, т.е. суть H_α состоит в том, что $\bar{x}_c \ne \bar{x}_н$. Примем уровень значимости α равным 0,05, тогда

$P(|\bar{x}_н - \bar{x}_c| \geq t\mu_{\bar{x}} = 0{,}05$ и критическая область соответственно задается неравенством $|\bar{x}_н - \bar{x}_c| \geq t\mu_{\bar{x}}$. По таблицам интегральной функции Лапласа определяем коэффициент доверия $t = 1{,}96$. Таким образом, величина предельного расхождения двух средних с вероятностью, равной 0,95, не должна превышать $t\mu_{\bar{x}} = 0{,}196$ (1,96×0,1). Следовательно, с вероятностью 0,95 доверительные пределы для генеральной средней при новом методе будут равны $2{,}64 < \bar{x}_н < 2{,}996$ $(\bar{x}_c \pm \Delta_{\bar{x}})$.

Средний расход материала при применении новой технологии составляет 2,6, т. е. попадает в критическую область. Следовательно, данные наблюдения не являются совместимыми с выдвинутой гипотезой H_0 о том, что между новой и существующей технологией производства изделий отсутствуют различия, значительные с точки зрения их влияния на материалоемкость.

2. Применение нового метода обработки приводит к снижению расхода сырья на единицу продукции, т.е. суть H_u состоит в том, что $\bar{x}_н < \bar{x}_c$. В этом случае рассматривается область больших отрицательных отклонений, т.е. при $\alpha = 0{,}05$ $P(\bar{x}_н < \bar{x}_c - t\mu_{\bar{x}}) = 0{,}05$. В данном варианте критическая область определяется неравенством. Нулевая гипотеза не будет опровергаться, если средний расход материала на единицу продукции будет больше величины 2,636 (2,8 — 0,1×1,64). Однако по новой технологии расход сырья составляет 2,6 усл. ед., и в этом случае с вероятностью 0,995 можно отвергнуть нулевую гипотезу, следовательно, применение новой технологии приводит к снижению расхода сырья на изготовление продукции.

Если гипотеза о величине центра распределения проверяется по результатам малой выборки, то следует учитывать, что отношение разности средних к стандартной ошибке выборки, т.е. величина $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\mu_{\bar{x}}}$ имеет распределение Стьюдента с $(n - 1)$ степенями свободы.

Допустим, например, что данные о расходе сырья на единицу продукции были получены по результатам проверки 26 изделий. Используем ранее приведенные данные о среднем расходе сырья при старом способе обработки ($\bar{x}_c = 2{,}6$ усл. ед.) и при новом способе обработки ($\bar{x}_н = 2{,}6$ усл. ед.). По результатам наблюдений значение эмпирического среднего квадратического отклонения составило 0,6. Расчетная величина t-критерия определяется по формуле:

$$t_{расч} = \frac{|\bar{x}_н - \bar{x}_c|}{\dfrac{S}{\sqrt{n-1}}} = \frac{|2{,}6 - 2{,}8| \cdot 5}{0{,}6} = 1{,}67.$$

По таблице распределения Стьюдента значение t-критерия для числа степеней свободы $(n - 1)$, равного 25, и уровня значимости 0,05 составляет 2,06 (см. Приложение 4). Так как расчетное значение не превышает табличное, то нулевая гипотеза не опровергается, т.е. у нас нет достаточных оснований считать, что применение новой технологии приводит к снижению материалоемкости.

6.12. Как оценить существенность расхождения двух выборочных средних или долей?

На следующем примере рассмотрим проверку гипотезы о существенности расхождения двух выборочных средних (для случая малой выборки).

По результатам оценки размеров дебиторской задолженности, проведенной финансовым директором фирмы, были внесены изменения в ее кредитную политику. По истечении отчетного квартала было решено проанализировать, действительно ли изменения в кредитной политике фирмы оказали влияние на сокращение срока оплаты дебиторской задолженности.

В таблице представлены данные о продолжительности сбора дебиторской задолженности (в днях) при старой и новой кредитной политике фирмы:

Тип кредитной политики фирмы	Продолжительность сбора дебиторской задолженности (в днях)									
Старая	39	42	28	35	38	30	30	29	36	34
Новая	28	31	29	30	24	37	27	22	33	31

Предположим, что фирма ставит задачу проверить при уровне значимости 5%, привело ли изменение ее кредитной политики к изменению продолжительности сбора дебиторской задолженности.

В данном случае нулевая гипотеза состоит в отсутствии различий в величине выборочных средних, т.е. H_0 реализуется при $\bar{x}_1 = \bar{x}_2$. Тогда альтернативная гипотеза H_u реализуется при $\bar{x}_1 \neq \bar{x}_2$. Уровень

значимости α для проверки выдвинутой гипотезы равен 0,05. Ответ на поставленный вопрос требует сопоставления разности двух выборочных средних с величиной средней квадратической ошибки этих средних, т.е. должна быть рассчитана по фактическим данным двух выборок величина t:

$$t_{\text{расч}} = \frac{\tilde{x}_1 - \tilde{x}_2}{\mu_{(\tilde{x}_1 - \tilde{x}_2)}}. \qquad (6.19)$$

При выдвинутой выше нулевой гипотезе предполагается, что обе выборки принадлежат к одной и той же генеральной совокупности со средней $\bar{x}$ и дисперсией σ^2.

Поскольку мы располагаем данными двух выборок и величиной двух дисперсий S_1^2 и S_2^2, оценку генеральной дисперсии следует определить как среднюю арифметическую взвешенную из выборочных дисперсий, а в качестве веса использовать число степеней свободы в каждой выборке:

$$\sigma^2 = \frac{\sum_{j=1}^{2} S_j^2 (n_j - 1)}{\sum_{j=1}^{2} (n_j - 1)}, \text{ или } \sigma^2 = \frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{n_1 + n_2 - 2}, \text{ а поскольку}$$

$$S_1^2 = \frac{\sum_{i=1}^{n_1} (x_{ij} - \tilde{x}_1)^2}{n_1 - 1} \text{ и } S_2^2 = \frac{\sum_{i=1}^{n_2} (x_{ij} - \tilde{x}_2)^2}{n_2 - 1}, \text{ то}$$

$$\sigma^2 = \frac{\sum_{i=1}^{n_1} (x_{i1} - \tilde{x}_1)^2 + \sum_{i=1}^{n_2} (x_{i2} - \tilde{x}_2)^2}{n_1 + n_2 - 2}$$ и формула стандартной ошибки

разности двух выборочных средних имеет вид:

$$\mu_{\tilde{x}_1 - \tilde{x}_2} = \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_1}} = \sqrt{\sigma^2 \frac{n_1 + n_2}{n_1 \cdot n_2}}. \qquad (6.20)$$

где σ^2 — оценка генеральной дисперсии по данным двух выборок;
n_1 и n_2 — количество наблюдений соответственно в первой и второй выборках.

Результаты расчетов для проверки выдвинутой гипотезы об отсутствии влияния изменений в кредитной политике фирмы на продолжительность сбора дебиторской задолженности приведены в табл. 6.10.

Таблица 6.10

Тип кредитной политики фирмы	Средняя продолжительность сбора дебиторской задолженности (в днях)	Число обследованных потребителей n_j	Сумма квадратов отклонений $\sum_{i=1}^{n_j} (x_{ij} - \tilde{x}_j)^2$
Старая	34,1	10	187,69
Новая	29,2	10	165,80
Итого		20	353,49

При предположении отсутствия существенных различий в продолжительности сбора дебиторской задолженности, т.е. верности нулевой гипотезы, величина $t_{\text{расч}}$ не должна превышать табличного значения t при данном уровне значимости и числе степеней свободы $(n_1 + n_2 - 2)$. По таблице распределения Стьюдента (см. Приложение 4) при уровне значимости $\alpha = 5\%$ и числе степеней свободы 18 (10 + 10 − 2) находим, что $t_{\text{расч}} = 2{,}101$. Вероятность получения по результатам расчетов значения $t_{\text{расч}}$, большего по абсолютной величине, чем 2,101, в условиях случайных расхождений выборочных характеристик составляет всего 0,05 и является событием практически невозможным.

Используя данные табл. 6.10, рассчитаем среднюю квадратическую ошибку разности двух выборочных средних:

$$\mu_{\tilde{x}_1 - \tilde{x}_2} = \sqrt{\frac{353{,}49}{10 + 10 - 2} \cdot \frac{10 + 10}{10 \cdot 10}} = 1{,}98 \text{ дня}.$$

Тогда

$$t_{\text{расч}} = \frac{\tilde{x}_1 - \tilde{x}_2}{\mu_{\tilde{x}_1 - \tilde{x}_2}} = \frac{34{,}1 - 29{,}2}{1{,}98} = 2{,}475.$$

Поскольку $t_{\text{расч}} = 2{,}475$ оказывается больше табличного при уровне значимости 5% ($t_{\text{расч}} = 2{,}101$), нулевая гипотеза не подтверждается, т.е. различие между средней продолжительностью сбора дебиторской задолженности при старой и новой кредитной политике фирмы нельзя объяснить случайностями выборки.

Выдвинутую нулевую гипотезу можно проанализировать иначе. Зная величину средней квадратической ошибки разности двух выборочных средних, можно с заданной вероятностью указать предел возможных расхождений двух выборочных средних $\Delta_{\tilde{x}_1 - \tilde{x}_2} = t\mu_{\tilde{x}_1 - \tilde{x}_2}$, где t — табличное значение критерия Стьюдента.

В рассматриваемом примере $\Delta_{\bar{x}_1-\bar{x}_2} = 2{,}101 \times 1{,}98 = 4{,}16$ дня. Поскольку фактическая разность двух средних составляет 4,90 дня (34,1 – 29,2), нулевая гипотеза не подтверждается. Таким образом, можно предположить, что изменение кредитной политики фирмы дало определенный результат и привело к снижению продолжительности сбора дебиторской задолженности.

Как проверить гипотезу о существенности расхождений двух выборочных долей, покажем на следующем примере. По результатам выборочного обследования домохозяйств двух областей Центрально-Черноземного района России в 1992 г. были получены данные о доле доходов от предпринимательской деятельности в источниках доходов домохозяйств.

Разность двух выборочных долей составляет 0,5% (7,4 – 6,9). Можно ли считать несущественными различия в доле доходов от предпринимательской деятельности в источниках доходов домохозяйств двух областей? Нулевая гипотеза заключается в том, что отсутствуют существенные различия в роли новых видов доходов, традиционных для рыночной экономики.

Тогда в качестве оценки генеральной доли будет использоваться средняя взвешенная из долей, полученных по результатам выборочных обследований каждой области:

$$P = \frac{\sum_{j=1}^{2} w_j n_j}{\sum_{j=1}^{2} n_j} = \frac{0{,}069 \cdot 14{,}2 + 0{,}074 \cdot 12{,}5}{14{,}2 + 12{,}5} = 0{,}0713,$$

т.е. оценка доли дохода от предпринимательской деятельности в генеральной совокупности составляет 7,13%. Ошибка разности двух долей при справедливости нулевой гипотезы будет рассчитываться по формуле:

$$\mu_{w_1-w_2} = \sqrt{\frac{P(1-P)}{n_1} + \frac{P(1-P)}{n_2}} = \sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} =$$

$$= \sqrt{0{,}0713 \cdot 0{,}9287\left(\frac{1}{14\,200} + \frac{1}{12\,500}\right)} = 0{,}00316.$$

Таким образом, средняя квадратическая ошибка разности двух выборочных долей составляет 0,316%.

Поскольку обе выборки достаточно большого объема, можно воспользоваться таблицей нормированной функции Лапласа

Таблица 6.11

Номер области	Количество обследованных домохозяйств (в тыс. шт.) n_j	Удельный вес доходов от предпринимательской деятельности (в %) W_j)
1	14,2	6,9
2	12,5	7,4

(см. Приложение 2) для определения значения коэффициента доверия t при вероятности 0,95 или 0,99. Соответствующие этим вероятностям значения t равны 1,96 и 2,58.

Расчетное значение t-критерия, равное отношению разности двух выборочных долей, составит:

$$t_{\text{расч}} = \frac{|10{,}069 - 0{,}0741|}{0{,}00316} = 0{,}631.$$

Поскольку расчетное значение t меньше табличного как при уровне значимости 5%, так и при уровне значимости 1%, мы не отвергаем нулевую гипотезу и делаем заключение о несущественности различий доли доходов от предпринимательской деятельности в источниках доходов домохозяйств двух областей.

Можно определить и максимально возможную величину расхождений двух выборочных долей с заданной вероятностью $\Delta_{w_1-w_2} = t\mu_{w_1-w_2}$.

При вероятности 0,95 $\Delta_{w_1-w_2} = 0{,}00619$ (1,96×0,00316).

При вероятности 0,99 $\Delta_{w_1-w_2} = 0{,}00815$ (2,58×0,00316).

Поскольку фактическая разность двух выборочных долей 0,005 меньше найденных предельных ошибок, можно принять нулевую гипотезу.

Таким образом, оба варианта проверки гипотезы о несущественности различий доли доходов от предпринимательской деятельности по результатам выборочных обследований в двух областях дали один и тот же результат.

Глава 7
КОРРЕЛЯЦИОННАЯ СВЯЗЬ И ЕЕ СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ

7.1. Что такое корреляционная связь и какой круг вопросов она изучает?

Статистические распределения характеризуются наличием более или менее значительной вариации в величине признака у отдельных единиц совокупности. Естественно, возникает вопрос о том, какие же причины формируют уровень признака в данной совокупности и каков конкретный вклад каждой из них. *Изучение зависимости вариации признака от окружающих условий составляет содержание теории корреляции*[13].

Изучение действительности показывает, что вариация каждого изучаемого признака находится в тесной связи и взаимодействии с вариацией других признаков, характеризующих исследуемую совокупность единиц. Вариация уровня производительности труда работников предприятий зависит от степени совершенства применяемого оборудования, технологии, организации производства, труда и управления и других самых различных факторов.

При изучении конкретных зависимостей одни признаки выступают в качестве факторов, обусловливающих изменение других признаков. Признаки этой первой группы в дальнейшем будем называть ***признаками-факторами*** **(*факторными признаками*)**, а признаки, которые являются результатом влияния этих факторов, будем называть ***результативными***. Например, при изучении зависимости между производительностью труда рабочих и энерговооруженностью их труда уровень производительности труда является результативным признаком, а энерговооруженность труда рабочих — факторным признаком.

Рассматривая зависимости между признаками, необходимо выделить прежде всего две категории связи: 1) функциональные; 2) корреляционные.

[13] Основоположниками теории корреляции считаются английские биометрики Ф. Гальтон (1822—1911) и К. Пирсон (1857—1936). Термин «корреляция» был заимствован из естествознания и обозначает соотношение, соответствие. Представление о корреляции как о взаимозависимости случайных переменных величин лежит в основе математико-статистической теории корреляции.

Функциональные связи характеризуются полным соответствием между изменением факторного признака и изменением результативной величины, и каждому значению признака-фактора соответствуют вполне определенные значения результативного признака.

В ***корреляционных связях*** между изменением факторного и результативного признака нет полного соответствия, воздействие отдельных факторов проявляется лишь в среднем при массовом наблюдении фактических данных. В простейшем случае применения корреляционной зависимости величина результативного признака рассматривается как следствие изменения только одного фактора (например, энерговооруженность труда рассматривается как причина роста производительности труда). Однако выделенный в данном примере в качестве основного признак-фактор не является единственной причиной изменения результативного признака, наряду с ним на величину результативного признака влияет множество других причин. Как уже указывалось, на формирование уровня производительности труда на предприятии более или менее существенное влияние оказывают факторы, характеризующие степень совершенства применяемой техники и технологии, уровень механизации и автоматизации труда, специализации производства, состав работающих, текучесть кадров и т.п.

Кроме того, сам признак-фактор, в свою очередь, может зависеть от изменения ряда обстоятельств. В сложном взаимодействии находится результативный признак — в более общем виде он выступает как фактор изменения других признаков. Отсюда результаты корреляционного анализа имеют значение в данной связи, а интерпретация этих результатов в более общем виде требует построения системы корреляционных связей.

Одновременное воздействие на изучаемый признак большого количества самых разнообразных факторов приводит к тому, что одному и тому же значению признака-фактора соответствует целое распределение значений результативного признака, поскольку в каждом конкретном случае прочие факторные признаки могут изменять силу и направленность своего воздействия.

При сравнении функциональных и корреляционных зависимостей следует иметь в виду, что при наличии функциональной зависимости между признаками можно, зная величину факторного признака, точно определить величину результативного признака. В отличие от жесткости однозначно функциональной связи корреляционные связи характеризуются множеством причин и следствий и с их помощью устанавливается тенденция изменения результативного признака при изменении величины факторного признака.

Необходимо отметить, что экономической теории принадлежит решающее слово в обосновании связей между теми или иными при-

знаками. При этом теоретический анализ должен показать, какие факторы влияют на исследуемый признак или же влияние каких факторов должно быть проверено. Статистическое выражение связи между явлениями может показать, что изменения одного из сопоставляемых признаков сопровождаются изменениями другого. Следовательно, нужно искать объяснение этим изменениям в их содержательном анализе. С помощью статистических методов изучения зависимостей можно установить, как проявляется теоретически возможная связь в конкретных условиях. Статистика не только отвечает на вопрос о реальном существовании намеченной теоретическим анализом связи, но и дает количественную характеристику этой зависимости. Зная характер зависимости одного явления от других, можно объяснить причины и размер изменений в явлении, а также планировать необходимые мероприятия для дальнейшего его изменения.

При исследовании корреляционных зависимостей между признаками решению подлежит широкий круг задач, к которым следует отнести: 1) предварительный анализ свойств моделируемой совокупности единиц; 2) установление факта наличия связи, определение ее направления и формы; 3) измерение степени тесноты связи между признаками; 4) построение регрессионной модели, т.е. нахождение аналитического выражения связи; 5) оценка адекватности модели, ее экономическая интерпретация и практическое использование.

7.2. Каковы основные предпосылки применения корреляционно-регрессионного метода анализа?

Для того чтобы результаты корреляционного анализа нашли практическое применение и дали желаемый результат, должны выполняться определенные требования в отношении отбора объекта исследования и признаков-факторов. Одним из важнейших условий правильного применения методов корреляционного анализа является требование однородности тех единиц, которые подвергаются изучению методами корреляционного анализа. Например, при корреляционном анализе зависимостей тех или иных технико-экономических показателей работы предприятий от определенных факторов должны быть отобраны предприятия, выпускающие однотипную продукцию, имеющие одинаковый характер технологического процесса и тип используемого оборудования, для предприятий добывающей промышленности определенную роль играет и географическое размещение предприятий.

При выполнении указанных общих требований далее необходимо провести ***количественную оценку*** однородности исследуемой совокупности по комплексу признаков. Одним из возможных вариантов такой оценки является расчет относительных показателей ***вариации***. Традиционно широкое распространение для этих целей получил ***коэффициент вариации***. Несколько реже применяется ***отношение размаха вариации к среднему квадратическому отклонению***. Вывод о неоднородности исследуемой совокупности по тому или иному признаку требует проверки гипотезы о принадлежности выделяющихся (аномальных) значений признака к исследуемой генеральной совокупности (см. гл. 6) или же группировки исходных данных с целью выделения более однородных групп.

Другим важным требованием, обеспечивающим надежность выводов корреляционного анализа, является требование ***достаточного числа наблюдений***. Как уже указывалось, влияние существенных причин может быть затушевано действием случайных факторов, «взаимопогашение» влияния которых на результативный показатель в известной мере происходит при выведении средней результативного показателя для массы случаев.

Определенные требования существуют и в отношении факторов, вводимых в исследование. Все множество факторов, оказывающих влияние на величину результативного показателя, в действительности не может быть введено в рассмотрение, да практически в этом и нет необходимости, так как их роль и значение в формировании величины результативного показателя могут иметь существенные различия. Поэтому при ограничении числа факторов, включаемых в изучение, наряду с качественным анализом целесообразно использовать и определенные количественные оценки, позволяющие конкретно охарактеризовать влияние факторов на результативный показатель (к оценкам можно отнести ***парные коэффициенты корреляции, ранговые коэффициенты*** при экспертной оценке влияния факторов и др.). Включаемые в исследование факторы должны быть независимыми друг от друга, так как наличие тесной связи между ними свидетельствует о том, что они характеризуют одни и те же стороны изучаемого явления и в значительной мере дублируют друг друга.

Все основные положения теории корреляции разрабатывались применительно к предположению о нормальном характере распределения исследуемых признаков. В этой связи целесообразным является изучение формы распределения, дающее возможность в известной мере обосновать правомерность применения методов корреляционного анализа. При значительном отклонении распределения признаков от нормального закона возникают проблемы с оценкой надежности рассчитанных по выборочным данным коэффициентов корреляции.

Следует отметить, что на практике часто сталкиваются с теми или иными отклонениями от исходных предпосылок. Однако это не означает, что мы должны отказываться от применения методов корреляционно-регрессионного анализа.

7.3. Какие приемы используются для выявления наличия корреляционной связи между двумя признаками?

Для ответа на вопрос о наличии или отсутствии корреляционной связи используется ряд специфических методов: так называемые элементарные приемы (параллельное сопоставление рядов значений результативного и факторного признаков, графическое изображение фактических данных с помощью поля корреляции, построение групповой и корреляционной таблиц), а также дисперсионный анализ. Простейшим приемом обнаружения связи является ***сопоставление двух параллельных рядов*** — ряда значений факторного признака и соответствующих ему значений результативного признака. Значения факторного признака располагают в возрастающем порядке и затем прослеживают направление изменения величины результативного признака. Результативный признак (функцию) в дальнейшем будем обозначать через y, а факторный признак — через x.

Например, по 20 туристическим фирмам были установлены затраты на рекламу (факторный признак) и количество туристов, воспользовавшихся услугами каждой фирмы (результативный признак)[14]. В табл. 7.1 фирмы ранжированы по величине затрат на рекламу.

Можно видеть, что в целом для всей совокупности фирм увеличение затрат на рекламу приводит к увеличению количества туристов, пользующихся услугами фирмы, хотя в отдельных случаях наличие такой зависимости может и не усматриваться. Например, сопоставим данные по фирмам с порядковыми номерами 7 и 11. Здесь мы видим даже обратное соотношение: у фирмы 11 количество туристов, воспользовавшихся ее услугами, меньше, чем у фирмы 7 и составляет 920 человек, хотя затраты на рекламу выше, чем у фирмы 7 на 1 усл. ден. ед. В каждом отдельном случае количество туристов, воспользовавшихся услугами фирмы, будет зависеть не только от размера затрат фирмы на рекламу, но и от того, как

[14] Поскольку речь идет об изложении методологии изучения взаимосвязей, мы ограничились совокупностью малого объема. Данные в примере условные.

Таблица 7.1

Ранжирование 20 туристических фирм по факторному и результативному признакам

Порядковые номера фирм	Затраты на рекламу (в усл. ден. ед.)	Количество туристов, воспользовавшихся услугами фирмы (человек)
1	8	800
2	8	850
3	8	720
4	9	850
5	9	800
6	9	880
7	9	950
8	9	820
9	10	900
10	10	1000
11	10	920
12	10	1060
13	10	950
14	11	900
15	11	1200
16	11	1150
17	11	1000
18	12	1200
19	12	1100
20	12	1000

сложатся прочие факторы, определяющие величину результативного признака.

В тех случаях, когда возрастание величины факторного признака влечет за собой возрастание и величины результативного признака, говорят о возможном наличии прямой корреляционной связи. Если же с увеличением факторного признака величина результативного признака имеет тенденцию к уменьшению, то можно предполагать обратную связь между признаками.

Однако наличие большого числа различных значений результативного признака, соответствующих одному и тому же значению признака-фактора, затрудняет восприятие таких параллельных рядов особенно при большом числе единиц, составляющих изучаемую совокупность. В таких случаях целесообразнее воспользоваться для установления факта наличия связи статистическими таблицами — корреляционными или групповыми.

Построение ***корреляционной таблицы*** начинают с группировки значений факторного и результативного признаков. Так как в при-

водимом примере факторный признак представлен всего пятью вариантами повторяющихся значений, достаточно в первом столбце табл. 7.2 записать эти варианты.

Для результативного признака необходимо определить величину интервала. Для этого воспользуемся формулой Стэрджеса:

$$h_y = \frac{y_{max} - y_{min}}{k} = \frac{1200 - 720}{5} = 96 \text{ человек},$$

где $k = 1+3{,}322 \lg 20$.

В корреляционной таблице факторный признак x, как правило, располагают в строках, а результативный признак y — в столбцах (графах) таблицы. Числа, расположенные на пересечении строк и столбцов таблицы, означают частоту повторения данного сочетания значений x и y (табл. 7.2).

Данная корреляционная таблица уже при общем знакомстве дает возможность выдвинуть предположение о наличии или отсутствии связи, а также выяснить ее направление. Если частоты в корреляционной таблице расположены по диагонали из левого верхнего угла в правый нижний угол (т.е. большим значениям фактора соответствуют большие значения функции), то можно предположить

Таблица 7.2

Группировка значений факторного и результативного признаков

Центральное значение интервала	768	865	962	1059	1156	f_x	$\bar{y}_j$
Группы по y / Группы по x	720–816	817–913	914–1010	1011–1107	1108–1207		
1	2	3	4	5	6	7	8
8	2	1				3	800
9	1	3	1			5	865
10		1	3	1		5	962
11		1	1	1	2	4	1035
12			1		1	3	1059
f_y	3	6	6	2	3	20	

Примечание. $\bar{y}_j$ — среднее значение результативного признака j-й группы значений факторного признака;
f_x — частота повторения данного варианта значения факторного признака для всей совокупности;
f_y — частота повторения результативного признака по всей совокупности.

наличие прямой корреляционной зависимости между признаками. Если же частоты расположены по диагонали справа налево (из правого верхнего угла в левый нижний угол), то предполагают наличие обратной связи между признаками.

Уместно подчеркнуть, что при рассмотрении корреляционной таблицы важно установить расположение основной доли частот. Возможны варианты, когда все клетки корреляционной таблицы окажутся заполненными. Однако это обстоятельство еще не означает, что корреляционная связь между признаками отсутствует. Нужно установить, как расположена в таблице основная масса частот. Для того чтобы сделать восприятие корреляционной таблицы более доступным и в целях более четкого выявления основной тенденции связи, можно для каждой строки рассчитать средние значения результативного признака, соответствующие определенному значению признака-фактора (каждая строка таблицы дает условное распределение y при определенном значении x).

Среднее число туристов для первой группы, состоящей из трех фирм, которые тратят на рекламу 8 усл. ден. ед., будет равно 800 человек:

$$\bar{y}_1 = \frac{768 \cdot 2 + 865 \cdot 1}{3} = 800 \text{ человек}.$$

Для следующей группы, состоящей из пяти фирм, у которых затраты на рекламу равны 9 усл. ден. ед., $\bar{y}_2 = (768 \cdot 1 + 865 \cdot 3 + 962 \cdot 1)/5 = 865$ человек и т.д. (рассчитанные таким образом средние представлены в графе 8 табл. 7.2).

Итак, увеличение средних значений результативного признака с увеличением значений факторного признака еще раз свидетельствует о возможном наличии прямой корреляционной зависимости числа туристов, воспользовавшихся услугами фирмы, от затрат фирмы на рекламу.

Корреляционная таблица позволяет компактно изложить материал, поэтому все последующие расчеты (показателей тесноты связи и параметров уравнения регрессии) можно вести по корреляционной таблице, что особенно удобно при значительном объеме исходных данных.

Другим возможным приемом обнаружения связи является построение ***групповой таблицы***. Все наблюдения разбиваются на группы в зависимости от величины признака-фактора, и по каждой группе вычисляются средние значения результативного признака (табл. 7.3).

Сравнив средние значения результативного признака по группам, можно сделать вывод, что рост затрат туристических фирм на рекламу влечет за собой увеличение числа клиентов, пользующих-

Таблица 7.3

Группы туристических фирм по затратам на рекламу (в усл. ден. ед.)	Число фирм в группе	Среднее число туристов, воспользовавшихся услугами данной группы фирм (человек)
1	2	3
8	3	790
9	5	860
10	5	966
11	4	1063
12	3	1100
Итого	20	

Примечание. Различие в величине среднего числа туристов каждой группы фирм, вычисленных в корреляционной (см. графу 8 табл. 7.2) и групповой таблицах (см. графу 3 табл. 7.3) объясняется тем, что при расчете средних в корреляционной таблице действительные значения результативного признака заменяются центральными значениями интервалов группировки.

ся услугами фирмы, т.е. в рассматриваемом примере можно предположить наличие прямой корреляционной зависимости между признаками.

Корреляционная зависимость отчетливо обнаруживается только при рассмотрении средних значений результативного признака, соответствующих определенным значениям факторного признака, так как при достаточно большом числе наблюдений в каждой группе влияние прочих случайных факторов при расчете групповой средней будет взаимопогашаться, и четче проявится зависимость результативного признака от фактора, положенного в основу группировки. Иными словами, предполагается, что все прочие причины, если они носят случайный характер, при определении средней по группам взаимопогашаются, т.е. дают в каждой группе один и тот же результат. Следовательно, различия в величине средних будут связаны только с различиями в величине данного факторного признака. Если бы связи между факторным и результативным признаками не было, то все групповые средние были бы приблизительно одинаковыми по величине[15].

Для предварительного выявления наличия связи и раскрытия ее характера, а в известной мере и для выбора формы связи, применяют также ***графический метод***. Используя данные об индивидуальных значениях признака-фактора и соответствующих ему значениях результативного признака (см. табл. 7.1), можно построить в прямоугольных координатах точечный график, который называют «полем корреляции». Для нашего примера «поле корреляции» имеет следующий вид, показанный на рис. 7.1.

Положение каждой точки на графике определяется величиной двух признаков — уровнем затрат на рекламу и соответствующим ему числом туристов, пользующихся услугами данной фирмы. Проведем на графике две оси, соответствующие средним значениям x и y. Тогда вся плоскость графика будет разделена на четыре части. Если бы точки, соответствующие значениям признака отдельных единиц, были равномерно распределены по всем четвертям графика, можно было предположить отсутствие связи между признаками. В нашем примере для большинства фирм можно видеть, что если затраты на рекламу ниже среднего ($x_i < \bar{x}$), то и число туристов ниже среднего ($y_i < \bar{y}$), а у тех фирм, которые тратят больше средних затрат на рекламу ($x_i > \bar{x}$), и $y_i > \bar{y}$ (точки правого верхнего квадрата). Всего лишь четыре точки отклоняются от названного соотношения (точки правого нижнего квадрата), но можно предполагать, что их влияние на степень тесноты связи не будет слишком

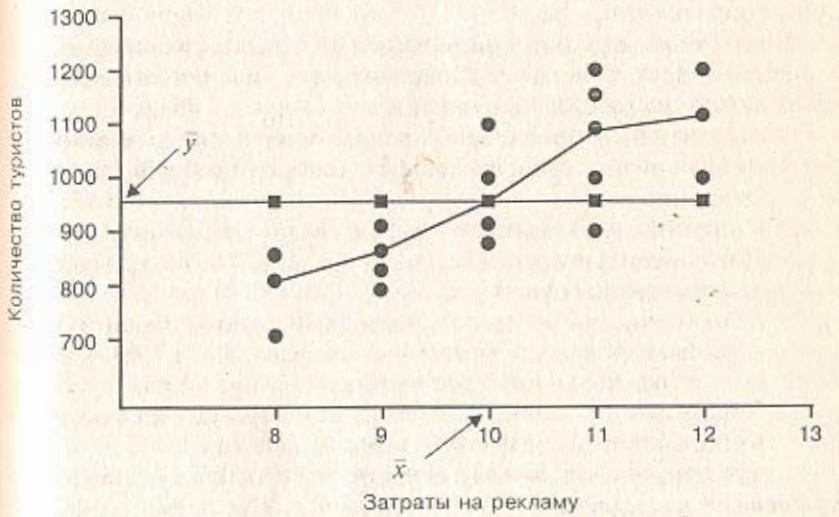


Рис. 7.1. Зависимость числа клиентов фирмы (y) от ее затрат на рекламу (x)

[15] Одним из недостатков аналитической группировки является неоднозначность результатов, которые зависят как от числа выделяемых групп, так и от установления границ интервалов.

значительным. В основном можно видеть, что точки корреляционного поля вытянуты определенной полосой вверх слева направо.

Имеющийся в нашем распоряжении статистический материал был сгруппирован (см. табл. 7.3) и по каждому значению затрат фирм на рекламу определены значения среднего числа туристов в группе (см. гр. 3 табл. 7.3). Нанеся эти средние на график и соединяя последовательно отрезками прямых соответствующие им точки, получим так называемую ***эмпирическую линию связи***.

Если эмпирическая линия связи по своему виду приближается к прямой линии, то можно предположить наличие прямолинейной корреляционной связи между признаками. Если же имеется тенденция неравномерного изменения значений результативного признака и эмпирическая линия связи будет приближаться к какой-либо кривой, то это может быть связано с наличием ***криволинейной корреляционной связи***.

7.4. Как измерить степень тесноты корреляционной связи в случае парной зависимости?

Показатели степени тесноты связи дают возможность охарактеризовать зависимость вариации результативного признака от вариации признака-фактора.

В известной мере они дополняют и развивают уже отмеченные приемы обнаружения связи. Показатели тесноты корреляционной связи можно использовать для решения следующих задач:

1) ответить на вопрос о необходимости изучения данной связи между признаками и целесообразности ее практического применения;
2) сопоставляя показатели тесноты связи для различных ситуаций, можно судить о степени различий в ее проявлении для конкретных условий;
3) и, наконец, сопоставляя показатели тесноты связи результативного признака с различными факторами, можно выявить те факторы, которые в конкретных условиях являются решающими и главным образом воздействуют на формирование величины результативного признака.

К простейшим показателям степени тесноты связи относят ***коэффициент корреляции знаков***, который был предложен немецким ученым Г. Фехнером (1801—1887). Этот показатель основан на оценке степени согласованности направлений отклонений индивидуальных значений факторного и результативного признаков от соответствующих средних. Для его расчета вычисляют средние значения результативного и факторного признаков, а затем проставляют знаки отклонений для всех значений взаимосвязанных пар признаков[16].

Обозначим через n_a число совпадений знаков отклонений индивидуальных величин от средней, через $n_н$ — число несовпадений знаков отклонений. Совпадения знаков отклонений будут иметь место у точек в левой нижней и правой верхней частях графика на рис. 7.1. В нижней правой четверти при $x_i \geq \bar{x}$ значения результативного признака ниже среднего ($y_i < \bar{y}$). Таким образом, в нашем примере $n_a = 16$ и $n_н = 4$. Формула коэффициента Фехнера записывается так:

$$K_ф = (n_a - n_н)/(n_a - n_н).$$

Коэффициент Фехнера может принимать различные значения в пределах от −1 до +1. Если знаки всех отклонений совпадут, то $n_н = 0$ и тогда коэффициент будет равен 1, что свидетельствует о возможном наличии прямой связи. Если же знаки всех отклонений будут разными, тогда $n_a = 0$ и коэффициент Фехнера будет равен −1, что дает основание предположить наличие обратной связи.

Рассчитаем его величину для рассматриваемого примера:

$$K_ф = (16 - 4)/(16 + 4) = 0{,}6.$$

Полученная величина коэффициента Фехнера свидетельствует о том, что можно предполагать наличие прямой зависимости между исследуемыми признаками.

Как видно из приведенной формулы для расчета коэффициента Фехнера, величина этого показателя не зависит от величины отклонений факторного и результативного признаков от соответствующей средней величины. Поэтому нельзя говорить о степени тесноты корреляционной связи, а тем более об оценке ее существенности на основании только коэффициента Фехнера. При малом объеме исходной информации коэффициент Фехнера практически решает ту же задачу, которая ставится при построении групповых и корреляционных таблиц, т.е. отвечает на вопрос о наличии и направлении корреляционной связи между признаками. В том случае, если построена корреляционная или же групповая таблица, дополнительный расчет коэффициента Фехнера не имеет практической ценности.

Более совершенным показателем степени тесноты связи является ***линейный коэффициент корреляции*** (r), который был предложен английским ученым К. Пирсоном.

[16] Средний размер затрат на рекламу 20 фирм составляет 9,95 усл. ден. ед., а среднее число туристов, воспользовавшихся услугами фирмы, — 952 человека.

При расчете этого показателя учитываются не только знаки отклонений индивидуальных значений признака от средней, но и сама величина таких отклонений, т.е. соответственно для факторного и результативного признаков определяют величины $(x_i - \bar{x})$ и $(y_i - \bar{y})$.

Сравнению могут подлежать отклонения, выраженные в относительных величинах, т.е. в долях среднего квадратического отклонения (их называют нормированными отклонениями). Так, для факторного признака будем иметь совокупность величин $t_{x_i} = \frac{x_i - \bar{x}}{\sigma_x}$, а для результативного $t_{y_i} = \frac{y_i - \bar{y}}{\sigma_y}$.

Полученные нормированные отклонения можно сравнивать между собой. Для того чтобы на основе сопоставления рассчитанных нормированных отклонений получить обобщающую характеристику степени тесноты связи между признаками для всей совокупности, рассчитывают среднюю величину произведений нормированных отклонений. Полученная таким образом средняя и будет являться линейным коэффициентом корреляции (r):

$$r = \frac{\sum_1^n t_{x_1} t_{y_1}}{n} = \frac{\sum_1^n \left(\frac{x_i - \bar{x}}{\sigma_x}\right)\left(\frac{y_i - \bar{y}}{\sigma_y}\right)}{n}, \qquad (7.2)$$

или, поскольку s_x и s_y для данных рядов являются постоянными, формула линейного коэффициента корреляции приобретает следующий вид:

$$r = \frac{\sum_1^n t_{x_1} t_{y_1}}{n} = \frac{\sum_1^n (x_i - \bar{x})(y_i - \bar{y})}{n\sigma_x\sigma_y}. \qquad (7.2а)$$

Вычисление коэффициента корреляции по формуле (7.2) является достаточно трудоемкой операцией. Выполнив несложные преобразования, можно получить следующую формулу для расчета линейного коэффициента корреляции:

$$r = \frac{n\sum_i x_i y_i - \sum_i x_i \sum_i y_i}{\sqrt{\left[n\sum_i x_i^2 - \left(\sum_i x_i^2\right)^2\right]\left[n\sum_i y_i^2 - \left(\sum_i y_i^2\right)^2\right]}}. \qquad (7.2б)$$

При пользовании этой формулой отпадает необходимость вычислять отклонения индивидуальных значений признаков от средней величины, что исключает ошибку в расчетах при округлении средних величин.

Линейный коэффициент корреляции может принимать любые значения в пределах от −1 до +1. Чем ближе коэффициент корреляции по абсолютной величине к 1, тем теснее связь между признаками. Знак при линейном коэффициенте корреляции указывает на направление связи: прямой зависимости соответствует знак «плюс», а обратной зависимости — знак «минус».

Если с увеличением значений факторного признака результативный признак имеет тенденцию к увеличению, то величина коэффициента корреляции будет находиться между 0 и 1. Если же с увеличением значений x результативный признак y имеет тенденцию к снижению, коэффициент корреляции может принимать значения в интервале от 0 до −1.

Используем данные табл. 7.1 и рассчитаем линейный коэффициент корреляции:

$$\sum_{i=1}^{20} x_i = 199; \quad \sum_{i=1}^{20} y_i = 19\,050; \quad \sum_{i=1}^{20} x_i y_i = 192;$$

$$\left(\sum_i x_i^2\right)^2 = 39\,601; \quad \left(\sum_i y_i^2\right)^2 = 362\,902\,500;$$

$$\sum_{i=1}^{20} x_i^2 = 2013; \quad \sum_{i=1}^{20} y_i^2 = 18\,497\,700;$$

$$r = \frac{20 \cdot 192\,310 - 199 \cdot 19\,050}{\sqrt{(20 \cdot 2013 - 39\,601)(20 \cdot 18\,497\,700 - 362\,902\,500)}} = 0{,}8105.$$

Полученная величина линейного коэффициента корреляции свидетельствует о возможном наличии достаточно тесной прямой зависимости между рассматриваемыми признаками.

Квадрат коэффициента корреляции (r^2) носит название ***коэффициент детерминации***. Для рассматриваемого примера его величина равна 0,6569, а это означает, что 65,69% вариации числа клиентов, воспользовавшихся услугами фирмы, объясняется вариацией затрат фирм на рекламу своих услуг.

В тех случаях, когда исходная информация представлена в виде корреляционной таблицы, нужно учитывать частоты повторений как индивидуальных значений факторного и результативного признаков, так и число повторений данного сочетания значений фактора и результата. При расчете линейного коэффициента корреляции по корреляционной таблице формула (7.2б) имеет вид:

$$r = \frac{n\sum_{i}^{n} x_i y_i f_{xy} - \sum_{i}^{n} x_i f_{xy} \sum_{i}^{n} y_i f_{xy}}{\sqrt{\left[n\sum_{i=1}^{n} x_i^2 f_x - \left(\sum_{i=1}^{n} x_i^2 f_x\right)^2\right]\left[n\sum_{i=1}^{n} y_i^2 f_x - \left(\sum_{i=1}^{n} y_i^2 f_x\right)^2\right]}}. \qquad (7.3)$$

Здесь еще раз следует напомнить, что сама по себе величина коэффициента корреляции не является доказательством наличия причинно-следственной связи между исследуемыми признаками, а является оценкой степени взаимной согласованности в изменениях двух признаков. Установлению причинно-следственной зависимости предшествует анализ качественной природы явлений. Но есть и еще одно обстоятельство, объясняющее формулировку выводов о возможном наличии связи по величине коэффициента корреляции. Связано это с тем, что оценка степени тесноты связи с помощью коэффициента корреляции производится, как правило, на основе более или менее ограниченной информации об изучаемом явлении. Возникает вопрос, насколько правомерно наше заключение по выборочным данным в отношении действительного наличия корреляционной связи в той генеральной совокупности, из которой была произведена выборка?

7.5. Как выполнить оценку существенности линейного коэффициента корреляции?

Принципиально возможны случаи, когда отклонение от нуля полученной величины выборочного коэффициента корреляции оказывается обусловленным неизбежными случайными колебаниями тех выборочных данных, на основании которых он вычислен. Особенно осторожно следует подходить к истолкованию полученных коэффициентов корреляции при незначительных объемах выборочной совокупности.

В этой связи и возникает необходимость оценки существенности линейного коэффициента корреляции, дающая возможность распространить выводы по результатам выборки на генеральную совокупность. В зависимости от объема выборочной совокупности и величины коэффициента корреляции предлагаются различные методы оценки его существенности. В отношении приводимых ниже критериев существенности можно сделать общее замечание, касающееся свойств исходной совокупности. Этим свойством является нормальное распределение значений признака в генеральной совокупности. Рассмотрим критерии, предлагаемые в статистической литературе.

Для малых значений коэффициента корреляции используется тот факт, что величина $t_{\text{расч}} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ при условии отсутствия связи в генеральной совокупности распределена по закону Стьюдента с $(n-2)$ степенями свободы.

Величина δ_r является средней квадратической ошибкой коэффициента корреляции:

$$\delta_r = \frac{\sqrt{1-r^2}}{\sqrt{n-2}}. \qquad (7.4)$$

Полученную величину $t_{\text{расч}}$ сравнивают с табличным значением t-критерия (число степеней свободы равно n — 2). Если рассчитанная величина превосходит табличное значение критерия t, то практически невероятно, что найденное значение обусловлено только случайными совпадениями x и y в выборке из генеральной совокупности, для которой действительное значение коэффициента корреляции равно нулю. Если же вычисленная величина $t_{\text{расч}}$ меньше, чем в таблице, то полагают, что коэффициент корреляции в генеральной совокупности в действительности равен нулю и соответственно эмпирический коэффициент корреляции существенно не отличается от нуля.

Применим указанный метод к оценке существенности корреляции между уровнем затрат туристических фирм на рекламу и числом туристов, воспользовавшихся услугами фирм. При объеме выборки, равном 20, и при условии, что величина коэффициента корреляции равна 0,8105:

$$t_{\text{расч}} = \frac{0{,}8105\sqrt{20-2}}{\sqrt{1-0{,}8105^2}} = \frac{0{,}8105 \cdot 4{,}243}{0{,}5857} = 5{,}872.$$

В таблице для числа степеней свободы $K = n - 2 = 18$ и уровня значимости 1% находим, что $t = 2{,}878$ (см. Приложение 4).

Таким образом, лишь с вероятностью меньше 1% можно утверждать, что величина $t = 5{,}872$ могла появиться в силу случайностей выборки. Такое событие является маловероятным, а потому можно считать с вероятностью 99%, что в генеральной совокупности действительно существует прямая зависимость между изучаемыми признаками, т.е. отличие выборочного коэффициента корреляции от нуля является существенным.

В тех случаях, когда линейный коэффициент корреляции, полученный по данным относительно малой выборки, близок к единице ($r > 0{,}8$), распределение его оценок отличается от нормального распределения или распределения Стьюдента, так как его величина ограничена единицей. В таких случаях для проверки его существенности рекомендуется использовать метод преобразованной корреляции, предложенный Р. Фишером.

Р. Фишер показал, что распределение логарифмической функции линейного коэффициента корреляции (обозначается эта функция через Z) приближается к нормальной кривой даже для выборок очень небольшого объема:

$$Z = \frac{1}{2} \ln \frac{1+r}{1-r}. \qquad (7.5)$$

Средняя квадратическая ошибка Z-распределения зависит только от объема выборки и определяется по формуле:

$$\sigma_Z = \frac{1}{\sqrt{n-3}}. \qquad (7.6)$$

По таблице соотношений между r и Z (см. Приложение 7), дающей возможность избежать вычислений логарифмов, находим, что коэффициенту корреляции 0,8105 соответствует $Z = 1{,}13$. Для числа наблюдений $n = 20$:

$$\sigma_Z = \frac{1}{\sqrt{20-3}} = \frac{1}{4{,}123} = 0{,}243.$$

Отношение Z к средней квадратической ошибке s_Z равно 4,65 (1,13 : 0,243). Поскольку отношение Z/σ_Z оказалось больше табличного значения критерия Стьюдента при уровне значимости 1%, можно полагать действительное наличие связи между признаками в генеральной совокупности.

7.6. Что такое эмпирическое корреляционное отношение и какова область его применения?

Коэффициент корреляции достаточно точно оценивает степень тесноты связи лишь в случае наличия линейной зависимости между признаками. При наличии же криволинейной зависимости линейный коэффициент корреляции недооценивает степень тесноты связи и даже может быть равен 0, а потому в таких случаях рекомендуется использовать в качестве показателя степени тесноты связи ***эмпирическое корреляционное отношение***. Этот показатель был предложен К. Пирсоном.

Расчет корреляционного отношения основан на использовании известной теоремы сложения дисперсий. Общая дисперсия результативного признака σ_0^2 может быть разложена на две составляющие. Первая составляющая — межгрупповая дисперсия δ^2 — характеризует ту часть колеблемости результативного признака, которая складывается под влиянием изменения признака-фактора, положенного в основу группировки:

$$\delta^2 = \frac{\sum_{j=1}^{k} (\bar{y}_j - \bar{y}_0) n_j}{\sum_{j=1}^{k} n_j},$$

где $\bar{y}_j$ — средние значения результативного признака в соответствующих группах;

$\bar{y}_0$ — общая средняя для всей совокупности;

n_j — число наблюдений в соответствующей группе;

k — число выделенных групп.

Вторая составляющая — средняя из внутригрупповых дисперсий $\overline{\sigma^2}$ — оценивает ту часть вариации результативного признака, которая обусловлена действием других, «случайных» причин:

$$\overline{\sigma^2} = \frac{\sum_{1}^{k} \sigma_j^2 n_j}{\sum_{1}^{k} n_j},$$

где σ_j^2 — дисперсия результативного признака в соответствующей группе.

Общая дисперсия равна: $\sigma_j^2 = \delta^2 + \overline{\sigma^2}$. Зная общую и межгрупповую дисперсии, можно оценить ту долю, которую составляет вариация под действием фактора x в общей вариации результативного признака y, т.е. найти отношение δ^2/σ^2.

Извлекая квадратный корень из этого отношения, мы получим эмпирическое корреляционное отношение (η):

$$\eta = \sqrt{\frac{\delta^2}{\sigma_0^2}}. \qquad (7.7)$$

Величина корреляционного отношения может быть рассчитана и по следующей формуле:

$$\eta = \sqrt{1 - \frac{\overline{\sigma^2}}{\sigma_0^2}}. \qquad (7.7а)$$

Величина корреляционного отношения будет равна 0, когда нет колеблемости в величине средних по выделенным группам. В тех случаях, когда внутригрупповая дисперсия близка к 0, т.е. практически вся вариация результативного признака обусловлена действием фактора x, величина корреляционного отношения близка к 1. Направление связи мы легко устанавливаем по данным групповой и корреляционной таблиц.

Воспользуемся данными групповой табл. 7.4 для расчета величины корреляционного отношения.

Используем итоговые данные графы 4 табл. 7.4 для расчета межгрупповой дисперсии:

$$\delta^2 = 236\,580/20 = 11\,829.$$

Среднее количество туристов, воспользовавшихся услугами одной фирмы, по всем 20 фирмам составило $\bar{y} = 952{,}5$ чел.

Величина общей дисперсии результативного показателя равна[17]:

$$\sigma_0^2 = \overline{y_i^2} - (\bar{y}_0)^2 = 18\,497\,700/20 - 907\,256{,}25 = 17\,628{,}75.$$

Таблица 7.4

Группы туристических фирм по затратам на рекламу (в усл. ден. ед.)	Число фирм в группе n_j	Среднее количество туристов, воспользовавшихся услугами данной группы фирм (человек)	$(\bar{y}_j - \bar{y}_0)^2 n_j$
1	2	3	4
8	3	790,0	79218,75
9	5	860,0	42781,25
10	5	966,0	911,25
11	4	1062,5	48400,00
2	3	1100,0	62268,75
Итого	20	952,5	236580,00

[17] Указанная сумма квадратов рассчитана по первичным данным, представленным в табл. 7.1.

Следовательно, величина корреляционного отношения, по данным приводимого примера, будет равна 0,819:

$$\eta = \sqrt{\frac{11\,829}{17\,628{,}75}} = 0{,}819.$$

Следует отметить, что вычисление корреляционного отношения возможно лишь при наличии достаточно большого числа данных, которые представлены в форме либо корреляционной, либо групповой таблицы. Вычисление корреляционного отношения при большом числе групп и малом числе наблюдений в каждой группе лишается смысла.

Определенный интерес представляет сопоставление величины линейного коэффициента корреляции и корреляционного отношения. Сравнив полученную величину корреляционного отношения с абсолютной величиной линейного коэффициента корреляции $|r| = 0{,}8105$, полученной при расчете по несгруппированным данным, можно видеть, что η незначительно больше r. Когда связь между переменными отклоняется от линейной формы, то η и r несколько отличаются по величине, причем η всегда больше r по абсолютной величине[18].

Сопоставление линейного коэффициента корреляции и эмпирического корреляционного отношения имеет смысл только в случае, если эти показатели вычислены для одинаковым образом сгруппированных данных, т.е. при сравнении и коэффициент корреляции, и корреляционное отношение должны быть вычислены либо по данным корреляционной таблицы, либо по первичным данным и групповой таблице, что предпочтительнее.

При проверке возможности использования линейной функции в качестве формы уравнения определяют разность квадратов $\eta^2 - r^2$, и если эта разность менее 0,1, то считается возможным применять линейное уравнение корреляционной зависимости. В нашем случае разность квадратов корреляционного отношения и линейного коэффициента корреляции равна 0,0139:

$$\eta^2 - \rho^2 = 0{,}819^2 - 0{,}8105^2 = 0{,}0139 (0{,}6708 - 0{,}6569),$$ что меньше 0,1.

Вывод о возможности использования линейной функции при моделировании зависимости между факторным и результативным признаками может быть сделан по критерию Z', рассчитываемому с учетом разности $\eta^2 - r^2$, или $Z' = 0{,}742\sqrt{n} \cdot \sqrt{\eta^2 - r^2}$. Зависимость может считается линейной, если $Z' < 2$.

[18] Доказательство этого положения приведено в кн.: Эдни Дж., Кендэл М. Дж. Теория статистики. М.: Госстатиздат, 1960. С. 300.

В нашем примере $Z' = 0{,}742\sqrt{20} \cdot \sqrt{0{,}0139} = 0{,}391$. По величине критерия можно сделать вывод о линейной форме связи между признаками.

Для проверки гипотезы о линейной зависимости более эффективно использование величины w^2:

$$w^2 = \frac{\eta^2 - r^2}{k-2} : \frac{1-\eta^2}{n-k}, \qquad (7.8)$$

которая подчиняется закону F-распределения с числом степеней свободы числителя $(k-2)$ и знаменателя $(n-k)$.

Задаваясь достаточно малым уровнем значимости (например, $\alpha = 0{,}05$), находим по таблицам F-распределения значение $F_{табл}$ при заданной величине α и соответствующем числе степеней свободы. Для рассматриваемого примера табличное значение F при 5%-ном уровне значимости и числе степеней свободы 3 $(k-2=5-2)$ и 15 $(n-k=20-5)$ равно 8,70 (Приложение 6).

Если w^2 окажется больше табличного значения F-критерия, то гипотезу о линейном виде регрессии можно считать статистически необоснованной.

В нашем примере:

$$w^2 = \frac{0{,}6708 - 0{,}6569}{5-2} : \frac{1-0{,}6708}{20-5} = 0{,}2105.$$

Сопоставление полученной величины с табличным значением F-критерия, равного 8,70, не позволяет отклонить гипотезу о линейности связи между затратами фирмы на рекламу и количеством туристов, воспользовавшихся ее услугами.

Применение линейного коэффициента корреляции для оценки степени тесноты связи между признаками, особенно в той части, которая связана с оценкой его существенности, является обоснованным лишь в условиях нормального или близкого к нормальному распределению признаков в изучаемой совокупности.

7.7. Какие показатели используют для измерения степени тесноты связи между качественными признаками?

Для определения величины линейного коэффициента корреляции необходимо знать численные значения факторного и результативного признаков. В некоторых же случаях мы можем встретиться с

такими качествами, которые не поддаются выражению числом единиц.

Эти обстоятельства заставляют прибегать к использованию так называемых *непараметрических методов*, позволяющих измерить интенсивность связи как между количественными признаками, форма распределения которых отличается от нормальной, так и между качественными признаками. В основу непараметрических методов положен принцип нумерации значений статистического ряда. Каждой единице совокупности присваивается порядковый номер в ряду, который будет упорядочен по уровню признака. Таким образом, ряд значений признака ранжируется, а номер каждой отдельной единицы будет ее рангом.

Можно получить предварительное представление о наличии или отсутствии связи между признаками, если сопоставить последовательность взаимного расположения рангов факторного и результативного признаков[19]. Для этого ранги индивидуальных значений факторного признака располагают в порядке возрастания, и если ранги результативного признака обнаруживают тенденцию к увеличению, можно предполагать наличие прямой связи; если же с увеличением рангов факторного признака ранги результативного признака уменьшаются, то это свидетельствует о возможном наличии между изучаемыми признаками образной связи.

Коэффициенты корреляции, основанные на использовании рангов, были предложены К. Спирмэном и М. Кендэлом. ***Коэффициент корреляции рангов Спирмэна*** (был использован им в начале XX в.) основан на рассмотрении разности рангов значений факторного и результативного признаков (обозначим ее d_i).

Формула коэффициента корреляции рангов Спирмэна, который обозначают через ρ, имеет следующий вид:

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}. \qquad (7.9)$$

Рассмотрим определение коэффициента Спирмэна на следующем примере.

Эксперты аналитического центра оценивали шансы депутатов на этапе предвыборной кампании в городскую Думу следующим образом (табл. 7.5 графы 1 и 2).

По результатам выборов можно проранжировать депутатов по числу поданных за них голосов избирателей (графа 3 табл. 7.5). Как

[19] При ранжировании значений факторного и результативного признаков следует использовать один и тот же принцип: либо от меньших значений к большим, либо от больших значений к меньшим.

Таблица 7.5

Порядковый номер кандидата	Ранг кандидатов по результатам оценки экспертов	Ранг депутата по числу поданных голосов на выборах	Разница рангов (гр.2—гр.3) $\lvert d_i \rvert$	d_i^2
1	2	3	4	5
1	7	5	2	4
2	4	6	2	4
3	1	2	1	1
4	3	7	4	16
5	10	8	2	4
6	5	3	2	4
7	9	10	1	1
8	2	1	1	1
9	8	9	1	1
10	6	4	2	4
Итого				40

видно из результатов сопоставлений рангов до выборов и после них, шансы депутатов не всегда оценивались экспертами с достаточной степенью точности, причем в одних случаях более точно, в других точность оценки была существенно ниже (например, у депутата под номером 4). Возникает вопрос: насколько точно результаты экспертной оценки предугадали действительные шансы депутатов быть избранными в городскую Думу?

Для ответа на поставленный вопрос рассчитаем коэффициент корреляции рангов Спирмэна, используя результаты расчетов в графе 5 табл. 7.5:

$$p = 1 - \frac{6 \cdot 40}{10(100 - 1)} = 0{,}758.$$

Поскольку коэффициенты корреляции рангов могут изменяться в пределах от −1 до +1 (как и линейный коэффициент корреляции), по результатам расчетов коэффициента Спирмэна можно предположить наличие достаточно тесной прямой зависимости между оценками экспертов на стадии предвыборной кампании и результатами выборов. Однако нельзя не учесть то обстоятельство, что ранговый коэффициент был рассчитан по небольшому объему исходной информации ($n = 10$). Не является ли отличие рангового коэффициента от нуля лишь результатом случайных совпадений оценок экспертов с результатами выборов по данным малого числа отобранных депутатов, можно ли распространить полученные выводы на генеральную совокупность?

Для совокупностей небольшого объема ($n < 30$) распределение рангового коэффициента корреляции не является нормальным, поэтому нецелесообразно использовать значения t по нормированной функции Лапласа для проверки гипотезы о величине рангового коэффициента корреляции. В Приложении 8 приводится таблица предельных значений коэффициентов корреляции рангов Спирмэна при условии верности нулевой гипотезы об отсутствии корреляционной связи при заданном уровне значимости и определенном объеме выборочных данных.

По таблице Приложения 8 находим, что при объеме выборки в 10 ед. и уровне значимости 5% критическая величина для рангового коэффициента корреляции составляет ±0,6364. Это означает, что вероятность получить величину коэффициента p, превышающую критическое значение при условии верности нулевой гипотезы (H_0: $p = 0$), будет менее 5%. В силу малой вероятности такое событие считается практически невозможным, и нулевая гипотеза может быть отвергнута.

Поскольку по результатам расчетов $p = 0{,}758$, что превышает критическую величину рангового коэффициента корреляции, можно принять альтернативную гипотезу о совпадении результатов выборов с оценками экспертов. Однако при уровне значимости $\alpha = 0{,}01$ критическое значение рангового коэффициента, которое может быть обусловлено случайными совпадениями рангов, составляет 0,7818 (см. Приложение 8). В этом случае с вероятностью 99% нулевая гипотеза не может быть отвергнута, т.е. величина $p = 0{,}758$ могла быть результатом случайных совпадений рангов обследованных депутатов, тогда как в генеральной совокупности связь между оценками экспертов и результатами выборов может отсутствовать. Поэтому общий вывод по результатам анализа может состоять в необходимости проведения расчетов по большему числу депутатов, а при отсутствии такой возможности следует относиться к оценкам экспертов данного аналитического центра с достаточной осторожностью.

М. Кендэл предложил еще одну меру связи между переменными x и y — ***коэффициент корреляции рангов Кендэла*** (τ):

$$\tau = \frac{2S}{n(n-1)}; \quad S = P + Q. \qquad (7.10)$$

Для вычисления τ надо упорядочить ряд рангов переменной x, приведя его к ряду натуральных чисел. Затем рассматривают последовательность рангов переменной y (табл. 7. 6).

Для нахождения суммы S (формула 7.10) находят два слагаемых P и Q. При определении слагаемого P нужно установить, сколько чисел, находящихся справа от каждого из элементов последовательности рангов переменной y, имеют величину ранга, превыша-

Таблица 7.6

Ранг депутатов по экспертной оценке *(x)*	1	2	3	4	5	6	7	8	9	10
Ранг депутатов по результатам выборов *(y)*	2	1	7	6	3	4	5	9	10	8

ющую ранг рассматриваемого элемента. Так, например, первому значению в последовательности рангов переменной y, т.е. числу 2, соответствует восемь чисел (7, 6, 3, 4, 5, 9, 10, 8), которые превышают ранг 2; второму значению соответствует также восемь чисел (7, 6, 3, 4, 5, 9, 10, 8), превышающих 1, и т.д. Суммируя полученные таким образом числа, мы получим слагаемое P, которое можно рассматривать как меру соответствия последовательности рангов переменной y последовательности рангов переменной x. Для нашего примера $P = 35$ (8 + 8 + 3 + 3 + 5 + 4 + 3 + 1).

Второе слагаемое Q характеризует степень несоответствия последовательности рангов переменной y последовательности рангов переменной x. Чтобы определить Q, подсчитаем, сколько чисел, находящихся справа от каждого из членов последовательности рангов переменной y, имеет ранг меньше, чем эта единица. Такие величины берутся со знаком «минус».

В рассматриваемом примере $Q = -10$ (−1−0−4−3−0−0−0−1−1). Следовательно, $S = P + Q = 35 - 10 = 25$.

Коэффициент корреляции рангов Кендэла в нашем примере равен:

$$\tau = (2 \times 25)/(10 \times 9) = 0{,}556.$$

Коэффициент Кендэла также изменяется в пределах от −1 до +1 и равен нулю при отсутствии связи между рядами рангов.

При достаточно большом числе наблюдений между коэффициентами корреляции рангов Спирмэна и коэффициентом корреляции рангов Кендэла существует следующее соотношение:

$$\rho = \frac{3}{2}\tau.$$

Существенность коэффициента корреляции рангов Кендэла проверяется при уровне значимости α по формуле:

$$\tau > t_\alpha \sqrt{\frac{2(2n+5)}{9n(n-1)}}. \qquad (7.11)$$

Могут встретиться случаи, когда невозможно установить ранговые различия нескольких смежных значений. В этих случаях принято брать средний ранг (даже если он будет дробным числом) и полученный средний ранг приписывать каждому из таких значений, т.е. переходить к матрице переформированных рангов. Например, двум факторам один из экспертов приписывает одинаковый ранг 3. Тогда каждому из факторов присваивается ранг 3,5, так как они поделили между собой третье и четвертое места (3 + 4)/2, а фактору, имевшему ранг 4, присваивается ранг 5 и т.д.

Если определяется теснота связи между k-м и l-м признаками, в рядах значений которых имеется соответственно q и g групп объединенных рангов, то формула коэффициента корреляции рангов Спирмэна примет вид:

$$\rho = \frac{\frac{n^3 - n}{6} - (T_k - T_l) - \sum_{i=1}^{n} d_i^2}{\sqrt{\left(\frac{n^3 - n}{6} - 2T_k\right)\left(\frac{n^3 - n}{6} - 2T_l\right)}}, \qquad (7.12)$$

где $T_k = \sum_{i=1}^{q} \frac{t_{k_i}^3 - t_{k_i}}{12}$; $T_l = \sum_{i=1}^{q} \frac{t_{l_i}^3 - t_{l_i}}{12}$;

t_{k_i} и t_{l_i} определяют количество единиц в i-й группе объединенных рангов соответствующего признака.

Скорректированная формула для вычисления коэффициента корреляции рангов Кендэла будет иметь вид:

$$\tau = \frac{S}{\sqrt{\left(\frac{n(n-1)}{2} - V_k\right)\left(\frac{n(n-1)}{2} - V_l\right)}}, \qquad (7.13)$$

где $V_k = \frac{\sum_{i=1}^{q} t_{k_i}(t_{k_i} - 1)}{2}$; $V_l = \frac{\sum_{i=1}^{q} t_{l_i}(t_{l_i} - 1)}{2}$.

При исследовании степени тесноты связи между качественными признаками, каждый из которых представлен в виде альтернативного признака, используют ***коэффициент ассоциации*** или ***коэффициент контингенции***[20]. Например, нужно оценить, влияют ли существующие формы повышения квалификации преподавателей университета на уровень их профессионального мастерства. Располагая данными о результатах аттестации студентами 320 преподавателей, из которых 240 повысили квалификацию, составляем табл. 7.8.

[20] Коэффициент ассоциации предложен английским статистиком Д. Юлом, а коэффициент контингенции — Д. Пирсоном.

Построенная в такой форме таблица носит название «таблицы четырех полей», частоты которой обозначим соответственно a, b, c, d.

Коэффициент ассоциации (K_A) определяется по формуле:

$$K_A = \frac{ad - bc}{ad + bc}. \qquad (7.14)$$

В приводимом примере его величина будет равна 0,22:

$$\left(\frac{163 \cdot 34 - 77 \cdot 46}{163 \cdot 34 + 77 \cdot 46} = +0{,}22\right).$$

Таким образом, по результатам проведенного в университете обследования вряд ли можно сделать убедительный вывод о повышении профессионального мастерства преподавателей в связи с повышением квалификации по одной из принятых форм (стажировка, факультет повышения квалификации, творческий отпуск и др.), поскольку степень тесноты связи невелика.

В тех случаях, когда хотя бы один из четырех показателей в «таблице четырех полей» отсутствует, величина коэффициента ассоциации будет равна единице, что дает преувеличенную оценку степени тесноты связи между признаками, и предпочтение следует отдать коэффициенту контингенции (K_K):

$$K_K = \frac{ad - bc}{\sqrt{(a+d)(b+d)(a+c)(c+d)}}. \qquad (7.15)$$

В качестве критерия наличия связи между качественными показателями с большим числом градаций можно использовать ***критерий «хи-квадрат»***. В табл. 7.9 приводятся результаты группировки предприятий отрасли по техническому и организационному уровню развития производства. Чисто визуально трудно ответить на вопрос о взаимосвязи между признаками, поэтому необходимо дать анализ распределения частот в таблице по строкам и графам.

Таблица 7.8

Группы преподавателей	Средний балл по сравнению с предыдущим по результатам аттестации		Всего
Повысившие квалификацию по одной из принятых форм	163 (a)	77 (b)	240
Не прошедшие повышение квалификации по принятым формам	46 (c)	34 (d)	80
Всего	209	111	320

Если признак, положенный в основу группировки по строкам таблицы (технический уровень), не зависит от признака, положенного в основу группировки по столбцам (организационный уровень), то в каждой строке (столбце) распределение частот должно быть пропорционально распределению их в итоговой строке (столбце). Такое распределение можно рассматривать в известной мере в качестве теоретического, частоты которого рассчитаны при предположении отсутствия связи между изучаемыми признаками.

Рассчитаем частоты внутри таблицы пропорционально распределению частот в итоговой строке (эти частоты приведены в табл. 7.9 в скобках):

$$f_{11} = \frac{25 \cdot 15}{53} = 7{,}1\,;\ f_{12} = \frac{25 \cdot 20}{53} = 9{,}4\,;\ f_{13} = \frac{25 \cdot 18}{53} = 8{,}5\,;$$

$$f_{21} = \frac{12 \cdot 15}{53} = 3{,}4\,;\ f_{22} = \frac{12 \cdot 20}{53} = 4{,}5\,;\ f_{23} = \frac{12 \cdot 18}{53} = 4{,}1 \quad \text{и т.д.}$$

Расчетное значение критерия «хи-квадрат» определяется по формуле:

$$\chi^2 = \sum_{i=1}^{k_2} \sum_{j=1}^{k_1} \frac{(f_{ij} - f_{ij}^1)^2}{f_{ij}^1}, \qquad (7.16)$$

где f_{ij} и f_{ij}^1 — соответственно эмпирические и теоретические частоты в i-й строке j-го столбца;

k_1 и k_2 — соответственно число групп в строках и столбцах таблиц.

Расчетное значение критерия по данным табл. 7.9 составит 5,064:

Таблица 7.9

Группировка предприятий отрасли по техническому и организационному уровням развития

Группы предприятий по техническому уровню	Группы предприятий по организационному уровню			Итого
	ниже среднего уровня	средний уровень	выше среднего уровня	
Ниже среднего уровня	7(7,1)	8(9,4)	10(8,5)	25
Средний уровень	4(3,4)	7(4,5)	1(4,1)	12
Выше среднего уровня	4(4,5)	5 (6,0)	7(5,5)	16
Итого	15	20	18	53

$$\frac{(7-7{,}1)^2}{7{,}1}+\frac{(8-9{,}4)^2}{9{,}4}+\frac{(10-8{,}5)^2}{8{,}5}+\frac{(4-3{,}4)^2}{3{,}4}+\frac{(7-4{,}5)^2}{4{,}5}+$$

$$+\frac{(1-4{,}1)^2}{4{,}1}+\frac{(4-4{,}1)^2}{4{,}1}+\frac{(5-6{,}0)^2}{6{,}0}+\frac{(7-5{,}5)^2}{5{,}5}=5{,}064.$$

По таблицам математической статистики устанавливается либо вероятность появления рассчитанного значения χ^2, соответствующего данному числу степеней свободы при предположении независимости признаков; либо табличное значение критерия «хи-квадрат», соответствующее уровню значимости 0,05 или 0,01. В рассматриваемом примере число степеней свободы равно четырем [$k = (3-1)(3-1) = 4$]. При уровне значимости 0,05 табличное значение χ^2 при $k = 4$ составляет 95 (см. Приложение 5), т.е. $\chi^2_{\text{расч}} < \chi^2_{\text{табл}}$. Следовательно, гипотеза об отсутствии влияния технического уровня развития на организационный уровень не опровергается при $\alpha = 0{,}05$.

На основе критерия «хи-квадрат» определяются показатели степени тесноты связи — коэффициенты взаимной сопряженности К. Пирсона и А. Чупрова. ***Коэффициент взаимной сопряженности К. Пирсона*** рассчитывается по формуле:

$$P=\sqrt{\frac{\chi^2}{n+\chi^2}}, \qquad (7.18)$$

где n — общее число наблюдений.

Коэффициент взаимной сопряженности А. Чупрова позволяет учесть число групп по каждому признаку и определяется следующим образом:

$$C=\sqrt{\frac{\chi^2}{n(k_1-1)(k_2-1)}}. \qquad (7.19)$$

По данным приводимого в табл. 7.9 примера получим следующие значения коэффициентов сопряженности:

Коэффициент К. Пирсона: $P=\sqrt{\frac{5{,}064}{53+5{,}064}}=0{,}295$.

Коэффициент А. Чупрова: $C=\sqrt{\frac{5{,}064}{53\sqrt{2\cdot 2}}}=0{,}219$.

Таким образом, между техническим и организационным уровнями развития отмечается некоторая зависимость, но степень тесноты связи невелика.

7.8. Как рассчитываются значение уравнения регрессии и параметры парной регрессии[21]?

Изучение корреляционных зависимостей основывается на исследовании таких связей между переменными, при которых значения одной переменной можно принять за зависимую переменную, которая «в среднем» изменяется в зависимости от того, какие значения принимает другая переменная, рассматриваемая как причина изменения зависимой переменной. Как уже было сказано, действие данной причины осуществляется в условиях сложного взаимодействия различных факторов, вследствие чего проявление закономерности затемняется влиянием случайностей. Вычисляя средние значения результативного признака для данной группы значений признака-фактора, мы отчасти элиминируем влияние случайностей. Вычисляя параметры теоретической линии связи, мы производим дальнейшее их элиминирование и получаем однозначное (по форме) изменение переменной y с изменением фактора x.

Теоретической линией регрессии называется линия, вокруг которой группируются точки корреляционного поля и которая указывает основное направление, основную тенденцию связи. Теоретическая линия регрессии должна отображать изменение средних величин результативного признака y по мере изменения величин факторного признака x при условии полного взаимопогашения всех прочих случайных по отношению к фактору x причин. Следовательно, эта линия должна быть проведена так, чтобы сумма отклонений точек поля корреляции от соответствующих точек теоретической линии регрессии равнялась нулю, а сумма квадратов этих отклонений была бы минимальной величиной.

Важным этапом регрессионного анализа является определение типа функции, с помощью которой характеризуется зависимость между признаками. Главным основанием для выбора вида уравнения должен служить содержательный анализ природы изучаемой зависимости, ее механизма. Вместе с тем теоретически обосновать форму связи каждого из факторов с результативным показателем можно далеко не всегда, поскольку исследуемые социально-экономические явления очень сложны и факторы, формирующие их уровень, тесно переплетаются и взаимодействуют друг с другом. Поэтому на основе теоретического анализа нередко могут быть сделаны самые общие выводы относительно направления связи, возможности его изменения в исследуемой совокупности, правомерности использования линейной зависимости, возможного наличия

[21] Термин «регрессия» был впервые использован Фрэнсисом Гальтоном в 1877 г.

экстремальных значений и т.п. Необходимым дополнением такого рода предположений должен быть анализ конкретных фактических данных.

Приблизительное представление о линии связи можно получить на основе эмпирической линии регрессии, или линии групповых средних (см. рис. 7.1). Эмпирическая линия обычно является ломаной, причем имеет более или менее значительный излом. Объясняется это тем, что влияние прочих неучтенных факторов, оказывающих воздействие на вариацию результативного признака, в средних погашается не полностью в силу недостаточно большого количества наблюдений, поэтому эмпирической линией связи для выбора и обоснования типа теоретической кривой можно воспользоваться при условии, что число наблюдений будет достаточно велико.

Можно также использовать опыт предыдущих исследований, и там, где выбранные формы уравнений связи давали удовлетворительный результат, рекомендовать их использовать в дальнейшем.

Одним из элементов конкретных исследований является сопоставление различных уравнений зависимости, основанное на использовании критериев качества аппроксимации эмпирических данных конкурирующими вариантами моделей. Наиболее часто для характеристики связей экономических показателей используют следующие типы функций:

линейную	$\tilde{y} = a + bx$;
гиперболическую	$\tilde{y} = a + b\frac{1}{x}$;
показательную	$\tilde{y} = ab^x$;
параболическую	$\tilde{y} = a + bx + cx^2$;
степенную	$\tilde{y} = ax^b$;
логарифмическую	$\tilde{y} = a + b\lg x$;
логистическую	$\tilde{y} = \frac{d}{1 + e^{a+bx}}$.

В рассматриваемом примере эмпирическая линия регрессии все же больше всего приближается к прямой и, следовательно, теоретическая линия регрессии может быть представлена уравнением вида:

$$\tilde{y} = a + bx.$$

Для нахождения параметров a и b уравнения регрессии используем метод наименьших квадратов. При применении метода наименьших квадратов для нахождения такой функции, которая наилучшим образом соответствует эмпирическим данным, считается, что сумма квадратов отклонений эмпирических точек теоретической линии регрессии должна быть величиной минимальной.

Критерий метода наименьших квадратов можно записать следующим образом:

$$S = \sum_{i=1}^{n} (y_i - \tilde{y}(x))^2 \to \min,$$

или, поскольку $\tilde{y} = a + bx$:

$$S = \sum_{i=1}^{n} [y_i - (a + bx)]^2 \to \min.$$

Следовательно, применение метода наименьших квадратов для определения параметров a и b прямой, наиболее соответствующей эмпирическим данным, сводится к задаче на экстремум.

Функция двух переменных $S(a,b)$ может достигнуть экстремума в том случае, когда первые частные производные этой функции равняются нулю, т.е. когда:

$$\frac{dS}{da} = 0 \quad \text{и} \quad \frac{dS}{db} = 0.$$

Вычисляя эти частные производные, получим:

$$\frac{dS}{da} = -2\sum_{1}^{n} (y_i - a - bx_i) = 0;$$

$$\frac{dS}{da} = -2\sum_{1}^{n} (y_i x_i - ax_i - bx_i^2) = 0.$$

После несложных преобразований получим систему нормальных уравнений способа наименьших квадратов для определения величины параметров a и b уравнения прямолинейной корреляционной связи по эмпирическим данным:

$$\begin{cases} \sum_{1}^{n} y_i = an + b\sum_{1}^{n} x_i \\ \sum_{1}^{n} y_i x_i = a\sum_{1}^{n} x_i + b\sum_{1}^{n} x_i \end{cases}. \qquad (7.20)$$

Используя результаты расчетов на с. 189 можно записать для нашего примера систему нормальных уравнений:

$$\begin{cases} 20a + 199b = 19\,050 \\ 199a + 2013b = 192\,310 \end{cases}.$$

Если данные сгруппированы (например, представлены в виде корреляционной таблицы), то система нормальных уравнений имеет вид:

$$\begin{cases} \sum_1^k y_i f_y = a\sum_1^k f_{xy} + b\sum_1^k x_i f_x \\ \sum_1^k y_i x_i f_{xy} = a\sum_1^k x_i f_x + b\sum_1^k x_i^2 f_x \end{cases},$$

где f_y — частота повторения данного варианта значения y;
f_x — частота повторения данного варианта значения x;
f_{xy} — частота повторения данного сочетания значений x и y.

Решая систему уравнений (7.20) относительно b, получим следующую формулу для определения этого параметра:

$$b = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n(\bar{x})^2}, \qquad (7.20а)$$

Значение параметра a получим, разделив обе части первого уравнения в формуле (7.20) на n:

$$a = \bar{y} - b\bar{x}. \qquad (7.20б)$$

В результате:

$a = 1183$, $b = 83{,}84$ и $\bar{y} = 1183 + 83{,}84x$.

Графическое изображение эмпирической и теоретической линии связи представлено на рис. 7.1.

7.9. Как проводится анализ параметров парной регрессии?

Параметр b в уравнении регрессии называют ***коэффициентом регрессии***. При наличии прямой корреляционной зависимости коэффициент регрессии имеет положительное значение, а в случае обратной зависимости коэффициент регрессии — отрицательный.

Коэффициент регрессии показывает, на сколько в среднем изменяется величина результативного признака y при изменении факторного признака x на единицу. Геометрически коэффициент регрессии представляет собой наклон прямой линии, изображающей уравнение корреляционной зависимости, относительно оси x (для уравнения $\bar{y} = a + bx$).

Коэффициент регрессии применяют для определения ***коэффициента эластичности***, который показывает, на сколько процентов изменится величина результативного признака y при изменении признака-фактора x на 1%.

Для определения коэффициента эластичности используется формула:

$$Э_x = b \cdot \frac{\bar{x}}{\bar{y}}. \qquad (7.21)$$

Для нашего примера коэффициент эластичности будет равен:

$$Э_x = 83{,}84 \cdot \frac{9{,}95}{952{,}5} = 0{,}8758.$$

Это означает, что при росте затрат фирмы на рекламу на 1% количество туристов, воспользовавшихся услугами фирмы, возрастет на 0,8758%.

Зная линейный коэффициент корреляции, оценивающий степень тесноты связи между изменениями факторного и результативного признаков, можно определить коэффициент регрессии в уравнении $\bar{y} = a + bx$ по следующей формуле:

$$b = r\frac{\sigma_y}{\sigma_x}, \qquad (7.22)$$

где σ_x и σ_y — средние квадратические отклонения соответственно значений результативного и факторного признаков.

Наличие этого соотношения дает возможность производить вычисление коэффициента корреляции и параметров уравнения линейной регрессии одновременно.

Воспользуемся данными примера, приведенного в табл. 7.1, для расчета параметров уравнения линейной зависимости.

Расчет показателей по несгруппированным данным привел к таким результатам:

$\sigma_x = 1{,}286$, $\sigma_y = 132{,}77$, $r = 0{,}8105$,

тогда $b = 0{,}8105\dfrac{132{,}77}{1{,}2836} = 83{,}83.$

Вопрос о целесообразности использования уравнения регрессии при решении задач обоснованного прогноза может быть проиллюстрирован на графике (рис. 7.2).

Значению факторного признака x_k соответствует по результатам наблюдения значение результативного признака y_k. Если воспользоваться уравнением регрессии для определения величины резуль-

тативного признака, получим величину $\bar{y}_k$. Общая величина отклонения наблюдаемого значения y_k от средней $\bar{y}$ будет тогда представлена на графике двумя отрезками. Отрезок 1 соответствует отклонению теоретического значения $\bar{y}_k$ от средней величины результативного признака ($\bar{y}_k - \bar{y}$) и представляет собой ту часть отклонения y от $\bar{y}$, которая может быть объяснена влиянием фактора x. Отрезок 2, соответствующий части общего отклонения ($y_k - \bar{y}_k$), характеризует остаточную часть общего отклонения и не может быть объяснен влиянием фактора x.

Использование уравнения регрессии приводит к получению лучших оценок для совокупности значений результативного признака по сравнению с ориентацией на общую среднюю, так как уменьшается величина отклонений:

$$\bar{y}_k - \bar{y} < y_k - \bar{y}.$$

Если рассмотреть общую сумму квадратов отклонений результативного признака от средней, т.е. величину $\sum_{i=1}^{n}(y_i - \bar{y}_i)^2$, сопоставить с ней меру «остаточной» вариации, т.е. величину $\sum_{i=1}^{n}(y_i - \bar{y}_i)^2$, то получится доля вариации, необъясняемая влиянием факторного признака x.

Тогда измерителем того, насколько велика степень связи между вариацией факторного и результативного признаков, будет величина:

$$1 - \frac{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}.$$

Извлекая квадратный корень из записанного выражения, получим показатель, называемый индексом корреляции:

$$i_{\bar{y}x} = \sqrt{1 - \frac{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}}. \qquad (7.23)$$

Для расчета индекса корреляции по приведенной формуле нужно предварительно для каждого значения факторного признака рассчитать по уравнению регрессии соответствующее ему значение результативного показателя. Поскольку такой порядок расчета весьма трудоемок, можно воспользоваться преобразованной формулой, учитывая, что для линейной зависимости:

$$\sum_{i=1}^{n}(y_i - \bar{y}_i)^2 = \sum_{i=1}^{n} y_i^2 - a\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i y_i,$$

а сумма квадратов отклонений может быть записана так:

$$\sum_{i=1}^{n}(y_i - \bar{y}_i)^2 = \sum_{i=1}^{n} y_i^2 - n(\bar{y})^2.$$

Тогда величина индекса корреляции может быть рассчитана по следующей формуле:

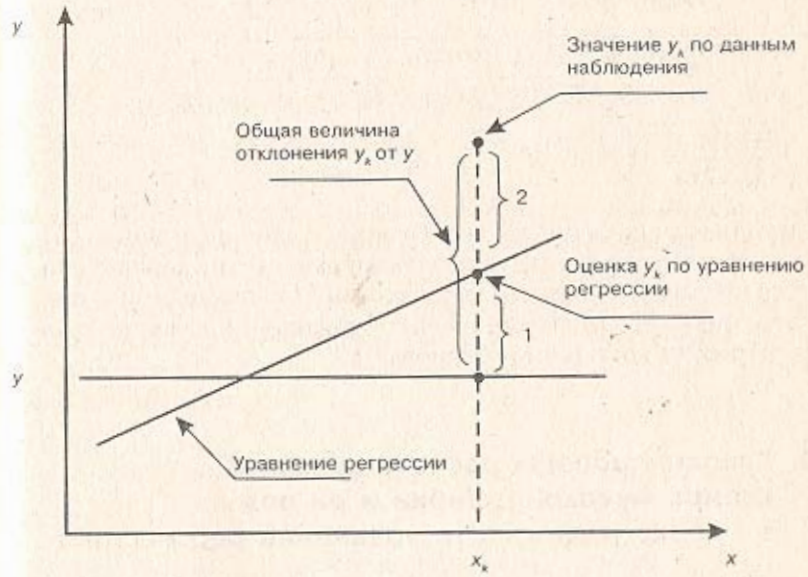


Рис. 7.2. Соотношение общей вариации признака, его составляющей, объясняемой влиянием фактора x, и остаточной вариации.

Отрезок 1 — величина ($\bar{y}_k - \bar{y}$); отрезок 2 — величина ($y_k - \bar{y}_k$).

$$i_{\bar{y}x} = \sqrt{\frac{a\sum_{i=1}^{n} y_i + b\sum_{i=1}^{n} x_i y_i - n(\bar{y})^2}{\sum_{i=1}^{n} y_i^2 - n(\bar{y})^2}}. \qquad (7.23а)$$

Из записи формул (7.23) и (7.23а) следует, что величина индекса корреляции зависит от того, какая форма уравнения регрессии была выбрана исследователем. А поэтому индекс корреляции в известной мере следует считать не показателем степени тесноты связи между изучаемыми признаками, а показателем степени близости выбранной теоретической линии регрессии к фактическим данным. Индекс корреляции может принимать значения в пределах от 0 до 1. Если $i_{\bar{y}x}$ равен или близок к нулю, это означает, что между переменными x и y либо нет связи, либо, если она и существует, то не может быть охарактеризована выбранной формой уравнения регрессии.

Для рассматриваемой зависимости числа туристов, воспользовавшихся услугами фирмы, и затратами фирмы на рекламу индекс корреляции составляет 0,8312:

$$i_{\bar{y}x} = \sqrt{\frac{118{,}3 \cdot 19\,150 + 83{,}84 \cdot 192\,310 - 20 \cdot (952{,}5)^2}{18\,497\,700 - 20 \cdot (952{,}5)^2}} =$$

$$= \sqrt{\frac{243\,590}{352\,575}} = 0{,}8312.$$

Близость величины индекса корреляции к единице в общем случае означает, что связь между признаками достаточно хорошо описывается избранным уравнением регрессии. Для приводимого примера это означает, что выбор линейной функции в качестве уравнения регрессии достаточно обоснован.

7.10. Каковы способы расчета средней квадратической ошибки и ее роль в оценке надежности уравнения регрессии?

Поскольку не все фактические значения результативного признака лежат на линии регрессии (см. рис. 7.1), более справедливо для записи уравнения корреляционной зависимости воспользоваться формулой $\bar{y} = a + bx + e$, где e отражает случайную составляющую вариации результативного признака, представленного на рис. 7.2 отрезком 2. Иногда рассеяние точек корреляционного поля настолько велико, что для принятия решений в управлении нет смысла пользоваться уравнением регрессии, так как погрешность в оценке анализируемого показателя будет чрезвычайно велика. Для всей совокупности наблюдаемых значений рассчитывается средняя квадратическая ошибка уравнения регрессии S_e, которая представляет собой среднее квадратическое отклонение фактических значений y_i относительно значений, рассчитанных по уравнению регрессии $\bar{y}_i$:

$$S_e = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2}{n - m}}, \qquad (7.24)$$

где S_e — средняя квадратическая ошибка уравнения регрессии;

y_i — фактические значения результативного признака, полученные по данным наблюдения;

$\bar{y}_i$ — значения результативного признака, рассчитанные по уравнению корреляционной связи и полученные путем подстановки значений факторного признака x в уравнение регрессии $\bar{y} = a + bx$;

m — число параметров в уравнении регрессии.

Заметим, что в данной формуле сумма квадратов отклонений y_i от $\bar{y}_i$ делится на число степеней свободы $(n - m)$, поскольку мы связали себя m степенями свободы в оценке теоретических значений результативного признака по уравнению регрессии с m параметрами. В случае линейного уравнения регрессии $m = 2$.

Используем формулу (7.15а) для упрощения расчета средней квадратической ошибки уравнения. В результате получим следующее выражение для определения величины S_e в случае линейной зависимости между x и y:

$$S_e = \sqrt{\frac{\sum_{i=1}^{n} y_i^2 - a\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i y_i}{n - 2}}, \qquad (7.24а)$$

где a и b являются параметрами уравнения регрессии.

Для рассматриваемого примера величина средней квадратической ошибки уравнения регрессии $\bar{y} = 118{,}3 + 83{,}84x$ будет равна 81,93 человека:

$$S_e = \sqrt{\frac{18\,497\,700 - 118{,}3 \cdot 19\,050 - 83{,}84 \cdot 192\,310}{20-2}} = 81{,}93 \text{ человека.}$$

Велика или мала полученная величина S_e? Можно сопоставить ее величину со средним значением результативного признака $\bar{y} = 952{,}5$ человека. Тогда $\frac{S_e}{y} \cdot 100\%$ будет составлять 8,60%. Можно также сравнить величину S_e со средним квадратическим отклонением результативного признака σ_y. Если $S_e < \sigma_y$, использование уравнения регрессии является целесообразным.

Чем меньше рассеяние эмпирических точек вокруг прямой, тем меньше средняя квадратическая ошибка уравнения. Таким образом, величина S_e служит показателем значимости и полезности прямой, выражающей соотношение между двумя признаками.

7.11. Как осуществить прогноз значений результативного признака по уравнению регрессии?

Средняя квадратическая ошибка уравнения дает нам возможность в каждом отдельном случае с определенной вероятностью указать, что величина результативного признака расположена в определенном интервале относительно значения, вычисленного по уравнению регрессии.

Покажем определение доверительных границ для результативного признака, т.е. тех границ, в пределах которых с заданной доверительной вероятностью будет находиться теоретическое значение y. Поскольку параметры уравнений регрессии определяются по выборочным данным, являясь функцией наблюдаемых значений, оценки параметров a и b содержат некоторую погрешность. Дисперсия значений результативного признака, определяемая по уравнению линейной зависимости, будет складываться из дисперсии параметра a и дисперсии параметра b.

Зная дисперсию показателя y и задаваясь уровнем доверительной вероятности, определяем доверительные границы результативного признака при значении факторного признака x_0 следующим образом:

$$\hat{y}_{x_0} - t_\alpha \frac{S_e}{\sqrt{n}} \sqrt{1 + \frac{(x_0 - \bar{x})^2}{\sigma_x^2}} \le Y \le \hat{y}_{x_0} + t_\alpha \frac{S_e}{\sqrt{n}} \sqrt{1 + \frac{(x_0 - \bar{x})^2}{\sigma_x^2}}, \qquad (7.25)$$

где t_α определяется в соответствии с уровнем значимости по распределению Стьюдента с $(n - m)$ степенями свободы.

Величина множителя $C_{x_0} = \sqrt{1 + \frac{(x_0 - \bar{x})^2}{\sigma_x^2}}$ будет вычисляться для каждого значения x_0. С удалением значения факторного признака от своего среднего арифметического значения величина C_{x_0} будет возрастать. Если представить доверительные границы графически, то они расположатся выше и ниже линии регрессии в виде ветвей гиперболы, ограничивая зону доверительных интервалов (рис. 7.3).

Покажем расчет доверительных границ числа туристов, которые воспользуются услугами фирмы, если затраты на рекламу будут составлять 10,5 усл. ден. ед.:

1) используя уравнение регрессии $\hat{y}_{x_0} = 118{,}3 + 83{,}84x$, находим число туристов, которые воспользуются услугами фирмы, если ее затраты на рекламу x_0 составят 10,5 усл. ден. ед.:

$$\hat{y}_{x_0} = 118{,}3 + 83{,}84 \times 10{,}5 = 998{,}6 \text{ человека:}$$

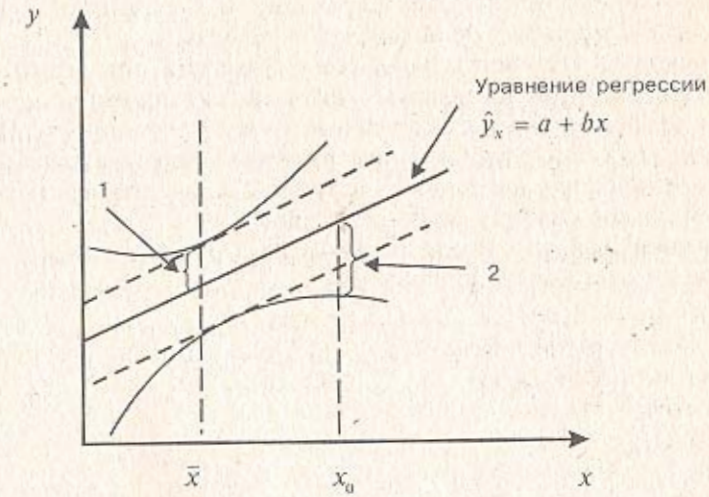


Рис. 7.3. Доверительные интервалы линейного уравнения регрессии

Отрезок 1 — $\frac{S_{\hat{y}}}{\sqrt{n}}$; *отрезок 2* — $\frac{S_{\hat{y}}}{\sqrt{n}} C_{x_0}$, *где* $C_{x_0} = \sqrt{1 + \frac{(x_0 - \bar{x})^2}{\sigma_x^2}}$

2) значение коэффициента t_α находим по таблице распределения Стьюдента для уровня значимости $\alpha = 0{,}05$ и числа степеней свободы, равного 18 (см. Приложение 4): $t_\alpha = 2{,}101$;

3) определяем множитель $C_{x_0} = \sqrt{1 + \frac{(10{,}5 - 9{,}95)^2}{1{,}2836^2}} = 1{,}1836$;

4) определяем величину $t_\alpha \frac{S_e}{\sqrt{n}} C_{x_0}$.

Рассчитанная величина S_e составила 81,93 человека, тогда

$t_\alpha \frac{S_e}{\sqrt{n}} C_{x_0} = 2{,}101 \frac{81{,}93}{\sqrt{20}} \cdot 1{,}1836 = 45{,}56$ человека.

Следовательно, с вероятностью 95% можно утверждать, что если фирма затратит на рекламу 10,5 усл. ден. ед., количество туристов, которые воспользуются услугами фирмы, будет находиться в интервале:

$998{,}6 - 45{,}6 \le y_{x_0} \le 998{,}6 + 45{,}6$, т.е. $953 \le y_{x_0} \le 1044$ человека.

Таким образом, учитывая «остаточную» вариацию результативного признака в данной совокупности фирм, можно полагать с вероятностью 95%, что число туристов, которые воспользуются услугами фирмы при затратах на рекламу 10,5 усл. ден. ед., будет не меньше 953 человек и не больше 1044 человек.

Поскольку параметры уравнения регрессии определяются по выборочным данным, являясь функцией наблюдаемых значений, оценки параметров a и b содержат некоторую погрешность. Поэтому, как и во всех случаях оценки параметров генеральной совокупности по выборочным данным, возникает задача проверки гипотезы о величине коэффициента регрессии.

Средняя квадратическая (стандартная) ошибка коэффициента регрессии b определяется по формуле:

$$S_b = \frac{S_e}{\sqrt{\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2}}, \qquad (7.26)$$

или

$$S_b = \frac{S_e}{\sigma_x \sqrt{n}}, \qquad (7.26а)$$

где S_e — средняя квадратическая ошибка уравнения регрессии;
σ_x — среднее квадратическое отклонение факторного признака;

S_b — средняя квадратическая ошибка коэффициента регрессии b.

Средняя квадратическая ошибка параметра a уравнения регрессии определяется по формуле:

$$S_a = \frac{S_e}{\sqrt{n}}, \qquad (7.27)$$

где S_a — средняя квадратическая ошибка параметра a.

Величину средней квадратической ошибки уравнения можно использовать и при выборе вида той или иной функции в качестве уравнения регрессии. Допустим, что в качестве уравнения регрессии, характеризующего зависимость числа туристов, воспользовавшихся услугами фирмы, от затрат фирмы на рекламу, мы избрали параболу второго порядка. Тогда $\bar{y}_x = a_0 + a_1 x + a_2 x^2$.

Параметры уравнения a_0, a_1 и a_2 находим путем решения следующей системы нормальных уравнений:

$$\sum_{i=1}^{n} y_i = a_0 n + a_1 \sum_{i=1}^{n} x_i + a_2 \sum_{i=1}^{n} x_i^2;$$

$$\sum_{i=1}^{n} y_i x_i = a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 + a_2 \sum_{i=1}^{n} x_i^3;$$

$$\sum_{i=1}^{n} y_i x_i^2 = a_0 \sum_{i=1}^{n} x_i^2 + a_1 \sum_{i=1}^{n} x_i^3 + a_2 \sum_{i=1}^{n} x_i^4.$$

Решив систему уравнений по данным приводимого примера, получим: $a_0 = -382$; $a_1 = 186$; $a_2 = -5{,}08$.

Уравнение параболы второго порядка примет следующий вид:

$$\bar{y}_x = -382 + 186x - 5{,}08\,x^2.$$

Величина средней квадратической ошибки параболы S_e определяется по формуле:

$$S_e = \sqrt{\frac{\sum_{i=1}^{n} y_i^2 - a_0 \sum_{i=1}^{n} y_i - a_1 \sum_{i=1}^{n} x_i y_i - a \sum_{i=1}^{n} x_i^2 y_i}{n - 3}}. \qquad (7.28)$$

В рассматриваемом примере $S_e = 83{,}15$ человека. Сравнив значения средних квадратических ошибок уравнений прямой и параболы, можно отдать предпочтение уравнению прямой, поскольку величина S_e для этого уравнения регрессии является меньшей.

При практическом использовании уравнений регрессии следует помнить, что экстраполяция допускается только тогда, когда существенно не изменяются условия формирования уровней призна-

ков, которые лежали в основе определения параметров уравнения регрессии. В противном случае использование уравнений для составления прогнозов должно быть отвергнуто. Необходим новый эмпирический материал, который отразит взаимосвязь признаков в новых условиях с определенными качественными сдвигами.

7.12. Как оценить роль отдельных факторов в формировании результативного показателя по уравнению множественной зависимости?

В приведенном примере рассматривалась зависимость между двумя признаками, т.е. речь шла о так называемой парной корреляции. На практике чаще всего изменение изучаемого признака зависит от действия нескольких причин. В таких случаях изучение корреляционной связи не может ограничиваться парными зависимостями, и в анализ необходимо включить другие признаки-факторы, существенно влияющие на изучаемую зависимую переменную. Одновременное изучение корреляции нескольких переменных проводится на основе использования методов множественной корреляции. Так, рассматривая уровень фондоотдачи на различных предприятиях одной отрасли, мы можем установить, что его величина зависит от размеров предприятия, структуры основных фондов, степени их изношенности, обновления и ряда других факторов.

Если обозначить факторы x_1, x_2, x_3,... x_m, то линейное уравнение множественной зависимости может быть записано так:

$$\hat{y}_{x_1 x_2 \ldots x_m} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \ldots + b_m x_m.$$

Рассчитав параметры уравнения множественной зависимости, определяем значение индекса корреляции по следующей формуле:

$$i = \sqrt{1 - \frac{S^2_{yx_j}}{\sigma^2_y}}, \qquad (7.23б)$$

где $S^2_{yx_j}$ — дисперсия эмпирических значений относительно значений, рассчитанных по уравнению регрессии, которая определяется путем деления остаточной суммы квадратов отклонений результативного признака на $(n - m - 1)$;

σ^2_y — дисперсия эмпирических значений результативного признака.

По параметрам полученного уравнения можно оценить роль каждого фактора в изменении уровня результативного показателя y. Это

может быть сделано путем прямой оценки по величине коэффициентов регрессии при каждом из факторов, а также по коэффициентам эластичности $Э_{x_j}$, стандартизированным частным коэффициентам регрессии β-коэффициентам и Δ-коэффициентам.

Коэффициенты уравнения множественной регрессии показывают абсолютный размер влияния факторов на уровень результативного показателя и характеризуют степень влияния каждого фактора на анализируемый показатель при фиксированном среднем уровне других факторов, входящих в модель.

Для сравнения оценок роли различных факторов в формировании моделируемого показателя следует дополнить абсолютные величины относительными. Так, частные коэффициенты эластичности показывают, на сколько процентов в среднем изменяется y с изменением признака-фактора x_j на 1% при фиксированном положении других факторов, и рассчитываются по формуле:

$$Э_j = a_j \frac{\bar{x}_j}{\bar{y}},$$

где a_j — коэффициент регрессии при j-м факторе.

β-коэффициенты показывают, на какую часть среднего квадратического отклонения (s_y) изменится зависимая переменная y с изменением соответствующего фактора x на величину своего среднеквадратического отклонения (s_j). Этот коэффициент позволяет сравнивать влияние колеблемости различных факторов на вариацию исследуемого показателя, на основе чего выявляются факторы, в развитии которых заложены наибольшие резервы изменения результативного показателя:

$$\beta_j = a_j \frac{\sigma_j}{\sigma_y}. \qquad (7.29)$$

Коэффициенты эластичности и β-коэффициенты взаимосвязаны следующим образом:

$$\beta_j = Э_j \frac{v_j}{v_y}, \qquad (7.29а)$$

где v_j — коэффициент вариации j-го факторного признака;

v_y — коэффициент вариации результативного признака.

Чтобы оценить долю влияния каждого фактора в суммарном влиянии факторов, включенных в уравнение регрессии, рассчитывают Δ-коэффициенты:

$$\Delta_j = \frac{r_{jy} \cdot \beta_j}{\sum_j r_{jy} \cdot \beta_j} = \frac{r_{jy} \cdot \beta_j}{R^2}. \qquad (7.30)$$

Содержательный анализ моделей в целях уточнения приоритетности факторов опирается на сравнение перечисленных коэффициентов. В этих целях, особенно при достаточно большом числе факторов, включаемых в уравнение регрессии, производится ранжирование факторов по величине коэффициентов эластичности, «бета» и «дельта».

Рассмотрим принципы анализа степени влияния факторов на примере, когда в уравнение регрессии были включены пять факторов (табл. 7.10).

Если сопоставить значения коэффициентов эластичности (графы 2 и 5 табл. 7.10), то можно видеть, что главным фактором изменения результативного показателя является фактор x_5; при его изменении на 1% y возрастает на 1,795%. Вторым по силе влияния на результативный показатель является фактор x_1 и т.д. (графа 5 табл. 7.10).

Сравнение значений β_j позволяет сделать вывод, что с учетом уровня колеблемости факторов наибольшие резервы в изменении результативного показателя заложены в увеличении факторов x_5, x_4 и x_1 (графа 6 табл. 7.10).

Сопоставление значений коэффициентов Δ_j позволяет сделать вывод, что наибольшую долю влияния имеет фактор x_5; роль этого фактора в вариации результативного показателя составляет 52,5% общего влияния пяти факторов на результативный показатель. Доля влияния других факторов значительно уступает доле влияния фактора x_5. Следовательно, наибольшие возможности в изменении результативного показателя связаны с изменением фактора x_5, затем x_1 и далее x_4.

Таблица 7.10

Факторы	Значения коэффициентов			Ранг факторов по величине коэффициентов			Средний ранг
	$Э_j$	β_j	Δ_j	$Э_j$	β_j	Δ_j	
1	2	3	4	5	6	7	8
x_1	0,173	0,204	0,162	2	3	2	2
x_2	0,133	0,114	0,076	4	5	5	5
x_3	0,108	0,144	0,104	5	4	4	4
x_4	0,158	0,253	0,133	3	2	3	3
x_5	1,795	0,732	0,525	1	1	1	1

При построении многофакторных корреляционных моделей одной из предпосылок обоснованности конечных результатов является требование возможно меньшей коррелированности включенных в модель признаков-факторов (отсутствие мультиколлинеарности). В случае линейной зависимости между факторами система нормальных уравнений не будет иметь однозначного решения, в результате чего коэффициенты регрессии и другие оценки окажутся неустойчивыми. Кроме того, наличие взаимосвязи факторов затрудняет экономическую интерпретацию уравнения связи, так как изменение одного из факторов влечет, как правило, изменение факторов, с ним связанных. Для исключения мультиколлинеарности было предложено несколько методов. На практике чаще всего используют чисто эмпирический подход к решению этой проблемы[22].

Рассмотрим матрицу парных коэффициентов корреляции (индекс 0 присвоен результативному признаку y), по которой можно оценить степень взаимной коррелированнсти признаков-факторов.

Факторы	y	x_1	x_2	...	x_j	...	x_m
y	1	r_{10}	r_{20}	...	r_{j0}	...	r_{m0}
x_1	r_{01}	1	r_{21}	...	r_{j1}	...	r_{m1}
x_2	r_{02}	r_{12}	1	...	r_{j2}	...	r_{m2}
...	...	...	...	...	...	...	...
x_j	r_{0j}	r_{1j}	r_{2j}	...	1	...	r_{mj}
...	...	...	...	...	...	...	...
x_m	r_{0m}	r_{1m}	r_{2m}	...	r_{jm}	...	1

В качестве критерия мультиколлинеарности может быть принято соблюдение следующих неравенств:

$$r_{x_j y} > r_{x_j x_k};$$

$$r_{x_k y} > r_{x_j x_k}.$$

Если приведенные неравенства (или хотя бы одно из них) не выполняются, то исключается тот фактор x_j или x_k, связь которого с результативным показателем y будет менее тесной. Вместе с тем

[22] Более эффективным средством решения сформулированной выше задачи исключения коллинеарных факторов является использование факторного анализа.

окончательный вывод о наличии или отсутствии мультиколлинеарности должен быть сделан в соответствии с экономическим содержанием и логикой взаимосвязи конкретных факторов.

Величина ***совокупного коэффициента корреляции*** по значениям парных коэффициентов может быть определена следующим образом:

$$R^2 = 1 - \frac{\begin{vmatrix} 1 & r_{10} & r_{20} & \cdot & r_{m0} \\ r_{10} & 1 & r_{21} & \cdot & r_{m1} \\ r_{20} & r_{12} & 1 & \cdot & r_{m2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{m0} & r_{1m} & r_{2m} & \cdot & 1 \end{vmatrix}}{\begin{vmatrix} 1 & r_{12} & \cdot & r_{m1} \\ r_{12} & 1 & \cdot & r_{m2} \\ r_{1m} & r_{2m} & \cdot & 1 \end{vmatrix}}. \qquad (7.31)$$

Величина R^2, называемая еще ***коэффициентом детерминации***, показывает, в какой мере вариация результативного признака обусловлена влиянием признаков-факторов, включенных в рассматриваемое уравнение корреляционной зависимости.

Величина совокупного коэффициента корреляции изменяется в пределах от 0 до 1 и численно не может быть меньше, чем любой из образующих его парных коэффициентов корреляции. Чем ближе совокупный коэффициент корреляции к единице, тем меньше роль неучтенных в модели факторов и тем больше оснований считать, что параметры регрессионной модели отражают степень эффективности включенных в нее факторов.

Для случая зависимости результативного признака от двух факторных признаков формула совокупного коэффициента корреляции имеет вид:

$$R_{y x_1 x_2} = \sqrt{\frac{r_{10}^2 + r_{20}^2 + 2 r_{10} r_{20} r_{12}}{1 - r_{12}^2}}. \qquad (7.31а)$$

Расчет параметров уравнения зависимости для такого варианта можно вести по формулам, в которых используются уже вычисленные значения парных коэффициентов корреляции и средних квадратических отклонений.

Так, для уравнения вида $\tilde{y}_{x_1 x_2} = b_0 + b_1 x_1 + b_2 x_2$, параметры b_0, b_1, b_2 могут быть найдены по следующим формулам:

$$b_1 = \frac{\sigma_y}{\sigma_1} \cdot \frac{r_{10} - r_{20} \cdot r_{12}}{1 - r_{12}^2};$$

$$b_1 = \frac{\sigma_y}{\sigma_2} \cdot \frac{r_{20} - r_{10} \cdot r_{12}}{1 - r_{12}^2}; \qquad (7.32)$$

$$b_0 = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2,$$

где σ_1, σ_2, σ_y — средние квадратические отклонения соответственно признаков x_1, x_2 и x_y;

r_{10}, r_{20}, r_{12} — парные коэффициенты корреляции.

Для более глубокого исследования связей между явлениями целесообразно установить степень тесноты связи между результативным признаком y и каждым из факторных признаков при исключении влияния других факторных признаков.

Для решения поставленной задачи определяют так называемые ***коэффициенты частной корреляции***, выявляющие степень «чистого» влияния факторного признака на результативный признак. Для расчета частных коэффициентов корреляции могут быть использованы парные коэффициенты корреляции.

Для случая зависимости y от двух признаков можно будет вычислить два коэффициента частной корреляции:

1) частный коэффициент корреляции $r_{01,2}$ между результативным признаком y и фактором x_2 при элиминировании фактора x_2 показывает, какую часть колеблемость y, вызванная фактором x_1, составляет в колеблемости y под действием всех факторов, кроме фактора x_1:

$$r_{01,2} = \frac{r_{10} - r_{20} \cdot r_{12}}{\sqrt{(1 - r_{12}^2)(1 - r_{20}^2)}}; \qquad (7.33а)$$

2) второй частный коэффициент корреляции $r_{02,1}$ характеризует зависимость результативного признака от фактора x_2, при исключении влияния фактора x_1:

$$r_{02,2} = \frac{r_{20} - r_{10} \cdot r_{12}}{\sqrt{(1 - r_{12}^2)(1 - r_{20}^2)}}. \qquad (7.33б)$$

Для общего случая частные коэффициенты корреляции можно определить таким образом:

$$r_{0m,1,2,\ldots,m-1} = \sqrt{\frac{R_m^2 - R_{m-1}^2}{1 - R_m^2}}, \qquad (7.33)$$

где R_m^2 — коэффициент детерминации результативного признака y с комплексом факторных признаков $x_1, x_2, \ldots, x_{m-1}, x_m$;

R_{m-1}^2 — коэффициент детерминации результативного признака y с комплексом признаков $x_1, x_2, \ldots, x_{m-1}$;

$r_{0m.1,2,\ldots,m-1}$ — частный коэффициент корреляции y с факторным признаком x_m при исключении влияния факторных признаков $x_1, x_2, \ldots, x_{m-1}$.

При малых значениях $r_{0m.1,2,\ldots,m-1}$ нет смысла вводить в уравнение $(m-1)$ фактор, так как эффективность нового уравнения регрессии, характеризующее зависимость от m факторов, возрастает незначительно.

Построение многофакторных регрессионных моделей позволяет дать количественное описание основных закономерностей изучаемых явлений, выделить существенные факторы, обусловливающие изменение экономических показателей, и оценить их влияние.

Полученные модели в основном используются в двух направлениях — для сравнительного анализа и в прогнозировании. Например, для выявления внутриотраслевых резервов повышения эффективности производства рассчитывается уравнение множественной зависимости, рассматриваемое в качестве экономико-статистической модели анализируемого показателя эффективности и характеризующее основные закономерности в формировании этого показателя для совокупности предприятий отрасли. На основе такого уравнения можно проанализировать и сравнить влияние каждого фактора на повышение эффективности в среднем по отрасли.

Разбив все предприятия на группы, различающиеся по условиям производства, определяемым организационно-техническим уровнем развития, можно с помощью модели соответствующего показателя эффективности сравнить результаты деятельности этих групп предприятий между собой и со средним отраслевым уровнем.

Процедура сравнительного анализа с использованием уравнений регрессии сводится к следующему. Для каждой группы предприятий, сформированных по организационно-техническому уровню развития, находятся не только средние значения показателей, характеризующих эффективность деятельности, но и показателей, определяющих уровень результативных признаков (под ними понимаются признаки-факторы, включенные в соответствующую регрессионную модель). Затем проводится сравнение средних значений факторных признаков сопоставляемых групп предприятий и определяются величины $\Delta\bar{x}_{ij}$, где j — номер признака-фактора; i — номер группы. Умножая разности $\Delta\bar{x}_{ij}$, найденные по сравниваемым группам, на величину коэффициентов регрессии b_j в многофакторной регрессионной модели, получаем эффект влияния различий в уровнях факторов по группам (табл. 7.11). Величина полученного эффекта ($\Delta\bar{x}_{ij} \times b_j$) позволяет судить о вкладе каждого фактора в изменение результативного показателя в данной группе по сравнению с другой группой. Рассмотрим по табл 7.11 схему сравнительного анализа, когда вся совокупность предприятий (единиц наблюдения) разбита на две группы.

Общая величина различий значений результативного показателя $\bar{y} = \bar{y}_1 - \bar{y}_2$ обусловлена изменением признаков-факторов, включенных в уравнение регрессии. Размер этого влияния определяется

Таблица 7.11

Общая схема сравнительного анализа с применением уравнений регрессии

Факторы	Средние значения факторов по группе		Разность средних значений факторов (гр.2 – гр.3)	Коэффициенты регрессий (b_j)	Эффект влияния на результативный показатель различий в средних уровнях факторов (гр.4 × гр.5)
	I группа	II группа			
1	2	3	4	5	6
x_1	$\bar{x}_{11}$	$\bar{x}_{12}$	$\Delta\bar{x}_1$	b_1	$b_1\Delta\bar{x}_1$
x_2	$\bar{x}_{21}$	$\bar{x}_{22}$	$\Delta\bar{x}_2$	b_2	$b_2\Delta\bar{x}_2$
x_3	$\bar{x}_{31}$	$\bar{x}_{32}$	$\Delta\bar{x}_3$	b_3	$b_3\Delta\bar{x}_3$
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
x_j	$\bar{x}_{j1}$	$\bar{x}_{j2}$	$\Delta\bar{x}_j$	b_j	$b_j\Delta\bar{x}_j$
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
x_m	$\bar{x}_{m1}$	$\bar{x}_{m2}$	$\Delta\bar{x}_m$	b_m	$b_m\Delta\bar{x}_m$
y	$\bar{y}_1$	$\bar{y}_2$	$\Delta\bar{y}$		$\sum_{j=1}^{m} b_j\Delta\bar{x}_j$

величиной $\sum_{j=1}^{m} b_j \Delta\bar{x}_j$. Разность $(\Delta\bar{y} - \sum_{j=1}^{m} b_j \Delta\bar{x}_j)$ может быть объяснена влиянием прочих неучтенных факторов. Чем выше доля влияния включенных в уравнение регрессии факторов, тем большую значимость приобретает использование уравнения регрессии в управлении деятельностью исследуемой совокупности предприятий. По результатам такого сравнительного анализа можно дать объективную оценку возможностей повышения эффективности производства и выявить внутрихозяйственные резервы в целом по отрасли.

Применение регрессионных моделей в прогнозировании требует осторожности в тех случаях, когда мы сталкиваемся с выходом фактических значений факторных признаков за границы значений признаков, на основе которых было рассчитано уравнение регрессии. Необходима также проверка возможности экстраполяции, связанная с оценкой неизменности общих условий формирования уровней признаков в совокупности.

Возможность широкого применения методов корреляционно-регрессионного анализа еще в недалеком прошлом сдерживалась высокой трудоемкостью необходимых расчетных процедур. Сегодня широкое распространение получили ***пакеты прикладных программ*** (ППП) по статистике, ликвидировавшие эти ограничения. Однако роль исследователя остается чрезвычайно важной как на этапах предварительной подготовки массива исходной информации, так и на этапе содержательной интерпретации полученных уравнений регрессии и их практического применения. Прежде всего исследователь обосновывает наличие причинной зависимости между признаками-факторами и результативным признаком. При предварительном анализе массива фактических данных важно оценить однородность совокупности исследуемых единиц, проверить возможность наличия выделяющихся наблюдений, обосновать принципы группировки единиц по значениям факторного признака. По результатам такого анализа формируется информационный массив для расчета показателей степени тесноты связи, а затем и параметров уравнения регрессии.

Целесообразны предварительная проверка гипотезы о линейности связи между факторным и результативным признаками до выбора математической формы уравнения регрессии и анализ матрицы парных коэффициентов корреляции. Все сказанное свидетельствует о необходимости выявления промежуточных этапов в последовательности обработки данных с помощью пакетов прикладных программ для внесения необходимых корректировок в процесс обработки данных.

Глава 8
РЯДЫ ДИНАМИКИ

8.1. Каковы виды рядов динамики?

Ряд расположенных в хронологической последовательности значений статистических показателей представляет собой временной (динамический) ряд. Каждый временной ряд состоит из двух элементов: во-первых, указываются моменты или периоды времени, к которым относятся статистические данные; во-вторых, приводятся те статистические показатели, которые характеризуют изучаемый объект в определенный момент или за указанный период времени.

Статистические показатели, характеризующие изучаемый объект, называют ***уровнями ряда***. Вид ряда динамики зависит не только от характера показателей, оценивающих изучаемый объект, но и от того, приводится ли показатель за какой-либо период или его величина показывается по состоянию на определенный момент времени. Статистические показатели, приводимые в динамическом ряду, могут быть абсолютными, относительными или средними величинами.

В качестве примера ряда динамики абсолютных величин рассмотрим показатели ввода в действие жилых домов в России. Каждый

Таблица 8.1

Ввод в действие жилых домов (в млн кв. м)

Годы	Всего построено	В том числе		
		государственными предприятиями и организациями	жилищно-строительными кооперативами	населением за свой счет и с помощью кредита
1	2	3	4	5
1990	61,7	48,4	2,9	6,0
1991	49,4	37,6	2,4	5,4
1992	41,5	29,2	2,1	4,9
1993	41,8	15,0	1,9	5,6
1994	38,5	10,7	1,8	7,0

Источник. Россия в цифрах. 1995. Краткий статистический сборник. Госкомстат России. М., 1995. С. 258.

уровень приведенного динамического ряда характеризует ввод в действие жилых домов за определенный промежуток времени — год. Такие ряды динамики называются ***интервальными***.

Отличительной особенностью интервальных рядов динамики абсолютных величин является возможность суммирования их уровней. Так же как и каждый уровень приведенного интервального ряда получается путем суммирования уровней за более короткие промежутки времени (месяцы, кварталы), сумма уровней за годы будет характеризовать ввод жилых домов за больший промежуток времени. В результате суммирования уровней интервального динамического ряда получаем так называемые ***накопленные итоги***, которые имеют реальное содержание.

Сумма же уровней моментного ряда динамики не имеет никакого реального содержания, и накопленные итоги для этих рядов не рассчитываются. Определенный смысл имеет расчет разностей уровней моментного динамического ряда, который будет характеризовать изменение уровня за определенный период времени. Так, зная, что численность населения России на начало 1998 г. и 1999 г. составила соответственно 146,74 млн и 146,34 млн человек, можно сказать, что за 1998 г. численность населения страны уменьшилась на 0,4 млн человек.

Ряды динамики относительных и средних величин состоят из производных статистических показателей, полученных в результате сопоставления между собой суммарных абсолютных данных.

Важнейшей проблемой построения динамических рядов является проблема сопоставимости уровней этих рядов, относящихся к различным периодам. Вследствие многих обстоятельств однородность величин, составляющих динамический ряд, может нарушаться, и таким образом нарушается сопоставимость уровней динамического ряда.

Покажем на примере необходимость учета требований сопоставимости уровней динамического ряда при подготовке информации для анализа изменений явлений во времени.

Рассмотрим данные о динамике (номинальной) средней месячной заработной платы одного работника в январе — июне 1999 г. (графа 2 табл. 8.2). По этим данным можно сделать вывод о тенденции роста этого показателя в течение всего рассматриваемого периода (графа 3 табл. 8.2).

Учитывая, что показатели средней заработной платы входят в систему показателей уровня жизни населения, можно ли сделать вывод о его повышении в 1999 г.? Вот здесь-то и возникает проблема сопоставимости доходов в денежном выражении, поскольку в условиях высокой инфляции (индексы потребительских цен приведены в графе 2 табл. 8.3) покупательная способность рубля будет

Таблица 8.2

Показатели динамики номинальной и реальной среднемесячной заработной платы одного работника за первое полугодие 1999 г.

Месяцы	Начисленная среднемесячная заработная плата одного работника (в руб.)	Темп роста к соответствующему периоду предшествующего года (в %)	
		номинальной среднемесячной заработной платы	реальной среднемесячной заработной платы
1	2	3	4
Январь	1167	115,5	58,6
Февраль	1199	120,0	59,0
Март	1385	130,3	62,8
Апрель	1423	136,4	64,0
Май	1472	141,1	65,1
Июнь	1619	145,1	65,7

Таблица 8.3

Показатели уровня жизни в первом полугодии 1999 г.

Месяцы	Темп роста к соответствующему периоду предшествующего года (в %)		Коэффициент опережения роста потребительских цен (гр. 2 : гр. 3)
	потребительских цен	начисленной среднемесячной заработной платы	
1	2	3	4
Январь	196,9	115,5	1,705
Февраль	203,3	120,0	1,694
Март	207,6	130,3	1,593
Апрель	213,1	136,4	1,562
Май	216,7	141,1	1,536
Июнь	220,7	145,1	1,521

Источник: Вопросы статистики. 1999. № 9. С. 56, 58.

снижаться. Поэтому наряду с номинальными денежными доходами при изучении уровня жизни следует исчислять реальные денежные доходы населения, на основе сопоставления которых можно анализировать тенденции в динамике уровня жизни. Так, используя данные графы 3 табл. 8.2, можно сказать, что в июне 1999 г. номи-

нальная среднемесячная заработная плата выросла на 45,1% по сравнению с июнем 1998 г. Однако реальный уровень среднемесячной заработной платы составил в июне 1999 г. 65,7% от уровня июня 1998 г., т.е. снизился на 34,6%. Реальную среднемесячную заработную плату получают путем деления начисленной ее величины на коэффициент роста (индекс) потребительских цен. Так, в январе 1999 г. уровень реальной средней месячной заработной платы одного работника составил 592,69 руб. $\left(\frac{1167}{196{,}9 : 100}\right)$, а ее изменение по сравнению с январем 1998 г. будет равно соотношению темпов роста номинальной заработной платы и потребительских цен. За январь это составит 58,6% $\left(\frac{115{,}5}{196{,}9} \cdot 100\%\right)$. Темпы роста реальной среднемесячной заработной платы приведены в графе 4 табл. 8.2.

С аналогичными проблемами исследователь сталкивается при анализе динамики макроэкономических показателей (валового внутреннего продукта, валового национального продукта) и показателей объема производства предприятий и отраслей, когда приходится учитывать влияние изменения цен, искажающее реальную динамику физических объемов производства.

Важное значение имеет использование единой методологии расчета уровней сравниваемых показателей на протяжении всего анализируемого периода. Особое значение эта проблема приобретает при международных сопоставлениях.

8.2. Каковы показатели изменения уровней временного ряда и методы их исчисления?

При изучении динамики необходимо решить целый ряд задач и осветить широкий круг вопросов, с тем чтобы охарактеризовать особенности и закономерности развития изучаемого объекта. К числу основных задач, возникающих при изучении динамических рядов, относятся следующие: 1) характеристика интенсивности отдельных изменений в уровнях ряда от периода к периоду или от даты к дате; 2) определение средних показателей временного ряда за тот или иной период; 3) выявление основных закономерностей динамики исследуемого явления на отдельных этапах и в целом за рассматриваемый период; 4) выявление факторов, обусловливающих изменение изучаемого объекта во времени; 5) прогноз развития явления.

Динамический ряд представляет собой ряд последовательных уровней, при сопоставлении которых между собой можно получить характеристику скорости и интенсивности развития явления. В результате сравнения уровней получается система абсолютных и относительных показателей динамики, к числу которых относятся абсолютный прирост, коэффициент роста, темп прироста, абсолютное значение одного процента прироста и пункты роста. Если сравнению подлежат несколько последовательных уровней, то возможны два варианта сопоставления:

1. Каждый уровень динамического ряда сравнивается с одним и тем же предшествующим уровнем, принятым за базу сравнения. В качестве базисного уровня выбирается либо начальный уровень динамического ряда, либо уровень, с которого начинается какой-то новый этап развития явления. Такое сравнение называется *сравнением с постоянной базой.*

2. Каждый уровень динамического ряда сравнивается непосредственно с предшествующим; такое сравнение называют *сравнением с переменной базой.*

Показатели динамики с постоянной базой (базисные показатели) характеризуют окончательный результат всех изменений в уровнях ряда от периода, к которому относится базисный уровень, до данного (i-го) периода. ***Показатели динамики с переменной базой (цепные показатели)*** характеризуют интенсивность изменения уровня от периода к периоду (или от даты к дате) в пределах изучаемого промежутка времени (рис. 8.1).

Абсолютный прирост (Δ_i) определяется как разность между двумя уровнями динамического ряда и показывает, на сколько данный уровень ряда превышает уровень, принятый за базу сравнения:

$$\Delta_i = y_i - y_0 , \qquad (8.1)$$

где Δ_i — абсолютный прирост;

y_i — уровень сравниваемого периода;

y_0 — уровень базисного периода.

При сравнении с переменной базой абсолютный прирост будет равен:

$$\Delta_i = y_i - y_{i-1} , \qquad (8.1а)$$

где y_{i-1} — уровень непосредственно предшествующего периода.

Абсолютный прирост с переменной базой иначе называют ***скоростью роста.***

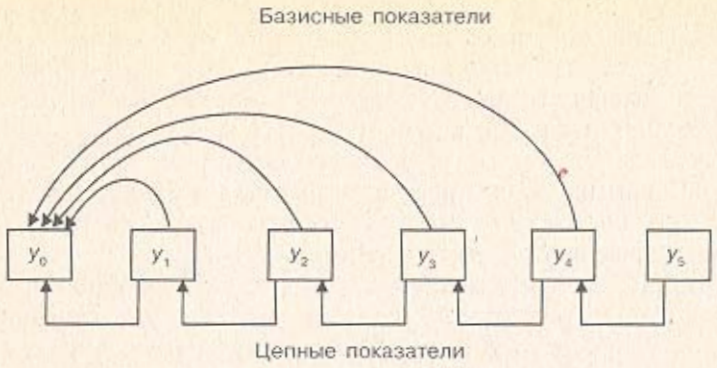


Рис. 8.1. Принципы построения цепных и базисных показателей динамики

Коэффициент роста определяется как отношение двух сравниваемых уровней и показывает, во сколько раз данный уровень превышает уровень базисного периода:

При сравнении с постоянной базой $K_i = \frac{y_i}{y_0}$. (8.2)

При сравнении с переменной базой $K_i = \frac{y_i}{y_{i-1}}$. (8.2а)

Если коэффициенты роста выражают в процентах, то их называют ***темпами роста***:

$$T_p = K \cdot 100\%. \quad (8.3)$$

Темп прироста показывает, на сколько процентов уровень данного периода больше (или меньше) базисного уровня. Этот показатель может быть рассчитан двояко:

1) как отношение абсолютного прироста к уровню, принятому за базу сравнения:

$$T_n = \frac{y_i - y_0}{y_0} \cdot 100\%, \quad (8.3а)$$

или

$$T_n = \frac{y_i - y_{i-1}}{y_{i-1}} \cdot 100\%; \quad (8.3б)$$

2) как разность между темпом роста (в процентах) и 100%:

$$T_n = T_p - 100\%. \quad (8.3в)$$

Рассмотрим расчет вышеуказанных показателей по ряду динамики ввода в действие жилых домов, построенных жилищно-строительными кооперативами в России, представленному в табл. 8.4 (см. данные табл. 8.1; графы 1 и 4).

Между показателями динамики, вычисленными с постоянной и переменной базой, существует определенная связь. Так, сумма абсолютных приростов с переменной базой дает общий прирост за исследуемый период:

$$\sum_{i=1}^{n-1} \Delta_i = y_n - y_1 = \Delta_{n/1},$$

где n — число уровней динамического ряда[23] ($-0,5 - 0,3 - 0,2 - 0,1 = -1,1$) (графы 3 и 4 табл. 8.4).

Таким же образом может быть осуществлен переход от коэффициентов роста с постоянной базой к коэффициентам роста, вычисленным с переменной базой и, наоборот. Если известны коэффициенты роста с постоянной базой, то частные от последовательного деления этих коэффициентов равны соответствующим коэффициентам роста с переменной базой:

$$K_{i/(i-1)} = \frac{K_{i/1}}{K_{(i-1)/1}}.$$

Таблица 8.4

Годы	Всего построено жилищно-строительными кооперативами (в млн кв. м)	Абсолютное изменение (в млн кв. м) по сравнению		Коэффициенты роста по сравнению		Темпы прироста (в %) по сравнению		Абсолютное значение 1% прироста (в тыс. кв. м)	Пункты роста (в %)
		с уровнем 1990 г.	с предшествующим годом	с 1990 г.	с предшествующим годом	с 1990 г.	с предшествующим годом		
1	2	3	4	5	6	7	8	9	10
1990	2,9	—	—	—	—	—	—	—	
1991	2,4	−0,5	−0,5	0,8276	0,8276	−17,24	−17,24	290	−17,24
1992	2,1	−0,8	−0,3	0,7241	0,8750	−27,59	−12,50	240	−10,35
1993	1,9	−1,0	−0,2	0,6552	0,9048	−34,48	−9,52	210	−6,89
1994	1,8	−1,1	−0,1	0,6207	0,9474	−37,93	−5,26	190	−3,45

[23] В подстрочных обозначениях стоят индексы уровней: в числителе — сравнительного, а в знаменателе — базы сравнения.

При наличии коэффициентов роста с переменной базой соответствующий базисный коэффициент роста находится путем перемножения цепных коэффициентов роста:

$$K_{i/1} = K_{2/1} \times K_{3/2} \times \ldots \times K_{(i-1)/(i-2)} \times K_{i/(i-1)};$$

(0,8276 × 0,8750 × 0,9048 × 0,9474 = 0,6207)
(графы 5 и 6 табл. 8.4).

При анализе относительных показателей динамики (темпов роста и темпов прироста) не следует рассматривать их изолированно от абсолютных показателей (уровней ряда и абсолютных приростов). Сравнение абсолютного прироста и темпа прироста за одни и те же периоды времени показывает, что замедление темпов прироста не всегда сопровождается уменьшением абсолютных приростов. Поэтому, чтобы правильно оценить значение полученного темпа прироста, его рассматривают в сопоставлении с показателем абсолютного прироста. Результат выражают показателем, который называют ***абсолютным значением одного процента прироста*** A_i:

$$A_i = \frac{y_i - y_{i-1}}{T_{ni/(i-1)}}, \qquad (8.4)$$

т.е. этот показатель рассчитывают как отношение абсолютного прироста к темпу прироста (в процентах) за тот же период времени.

Если преобразовать эту формулу, то получим следующее выражение:

$$A_i = \frac{y_i - y_{i-1}}{\frac{y_i - y_{i-1}}{y_{i-1}} \cdot 100\%} = \frac{y_{i-1}}{100} = 0{,}01 y_{i-1}. \qquad (8.4а)$$

Абсолютные значения одного процента прироста представлены в графе 9 табл. 8.4.

В тех случаях, когда сравнение производится с достаточно отдаленным периодом времени, принятым за базу сравнения, рассчитывают так называемые пункты роста, которые представляют собой выраженную в процентах разность базисных темпов роста (прироста), двух смежных периодов (см. графу 10 табл. 8.4). Так, для 1994 г. пункт роста составит −3,45 = −37,93 — (−34,48) (см. графу 7 табл. 8.4).

В отличие от темпов прироста, которые нельзя ни суммировать, ни перемножать, пункты роста можно складывать, в результате чего получаем темп прироста соответствующего периода по сравнению с базисным периодом, т.е. по данным нашего примера сумма пунктов роста равна −37,93, что соответствует темпу прироста уровня 1994 г. по сравнению с 1990 г. (−17,24 −10,35 − 6,89 −3,45 = −37,93) (см. графу 7 табл. 8.4).

При сопоставлении динамики развития двух явлений можно использовать показатели, представляющие собой отношения темпов роста или темпов прироста за одинаковые отрезки времени по двум динамическим рядам. Эти показатели называют ***коэффициентами опережения***:

$$K_{оп} = \frac{T'_р}{T''_р}, \text{ или } K_{оп} = \frac{T'_п}{T''_п},$$

где $T'_р$, $T''_р$, $T'_п$, $T''_п$ — соответственно темпы роста и темпы прироста сравниваемых динамических рядов.

С помощью этих коэффициентов могут сравниваться динамические ряды одинакового содержания, но относящиеся или к разным территориям (странам, регионам, районам и т.п.), или к различным организациям (министерствам, предприятиям), а также ряды разного содержания, характеризующие один и тот же объект. Например, по данным табл. 8.3 рассчитаем коэффициенты опережения роста потребительских цен по сравнению с ростом начисленной средней месячной заработной платы в январе—июне 1999 г. (см. графу 4). Данные этой таблицы показывают, что рост потребительских цен за этот период опережал рост номинальной средней месячной заработной платы в 1,521—1,705 раза, что не могло не сказаться на снижении уровня реальной заработной платы.

Коэффициенты опережения, рассчитанные разными способами, различны по величине, однако тенденция, которая отображена в их изменении, одинакова: за январь—июнь 1999 г. рост потребительских цен опережал рост средней месячной заработной платы, хотя размеры этого опережения имели тенденцию к снижению.

8.3. Каковы средние характеристики ряда динамики?

Для обобщающей характеристики динамики исследуемого явления за ряд периодов определяют различного рода средние показатели. Рассмотрим две категории этих показателей:

1) средние уровни ряда;

2) средние показатели изменения уровней ряда.

Метод расчета ***среднего уровня*** динамического ряда зависит от вида временного ряда.

Для ***интервального динамического ряда*** абсолютных показателей средний уровень за период определяется по формуле простой средней арифметической:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}, \qquad (8.6)$$

где n — число уровней ряда.

Средний уровень ***моментного динамического ряда*** определяется несколько иначе. Рассмотрим вычисление среднего уровня для такого моментного динамического ряда, когда промежутки между датами одинаковы. Например, определим средний месячный остаток материалов на складе предприятия в течение первого квартала текущего года, если известно, что остаток материалов на складе на 1 января составил 242 млн руб., на 1 февраля — 251 млн, на 1 апреля — 186 млн руб. При вычислении среднего уровня моментного ряда условно предполагается непрерывное, равномерное изменение уровня в промежутках между двумя датами. Основываясь на этом предположении, определим средние остатки материалов на складе за каждый месяц как половину суммы остатков на начало и конец месяца. Средние остатки за месяц соответственно будут равны:

за январь — 246,5 млн руб. $\left(\frac{242+251}{2}\right)$; за февраль — 232,0 млн. руб. $\left(\frac{251+213}{2}\right)$; за март — 192,5 млн руб. $\left(\frac{213+186}{2}\right)$.

Средний остаток за квартал определяется как простая средняя арифметическая $\bar{y} = \frac{246{,}5 + 232{,}0 + 192{,}5}{3} = 229$ млн руб.

Приведенный расчет среднего уровня можно представить формулой средней хронологической:

$$\bar{y}_{хр} = \frac{\frac{y_1}{2} + y_2 + y_3 + \ldots + y_{n-1} + \frac{y_n}{2}}{n-1}, \qquad (8.7)$$

где n — число дат;

$y_1, y_2, \ldots, y_n$ — уровни ряда в последовательные моменты времени.

В нашем примере средние остатки на складе будут равны:

$$\bar{y}_{хр} = \frac{\frac{242}{2} + 251 + 213 + \ldots + \frac{186}{2}}{4-1} = 229 \text{ млн руб.}$$

Для определения среднего уровня моментного ряда с неравными промежутками между временными датами вычисляется средняя арифметическая взвешенная; в качестве весов принимается продолжительность промежутков времени между моментами, в которые происходят изменения в уровнях динамического ряда:

$$\bar{y}_{хр} = \frac{\sum_{i=1}^{n} y_i t_i}{\sum_{i=1}^{n} t_i}, \qquad (8.8)$$

где t_i — количество дней (месяцев) между смежными датами.

Например, на 1 января отчетного года стоимость основных средств предприятий составляла 75 млрд руб. В марте были приобретены основные средства на сумму 2 млрд руб., в мае выбыло основных средств на 7 млрд руб., а в сентябре было приобретено еще основных средств на 8 млрд руб. Определим среднюю годовую стоимость основных средств предприятия. Для удобства расчетов данные представим в виде табл. 8.5.

Среднегодовая стоимость основных средств предприятия по данным приводимого примера составит 74,417 млрд руб.:

$$\bar{y} = \frac{893}{12} = 74{,}417 \text{ млн руб.}$$

Если же ориентироваться на стоимость основных средств предприятия на начало и конец отчетного года, т.е., например, использовать показатели балансового отчета предприятия, то получим иной показатель среднегодовой стоимости основных средств предприятия:

$$\bar{y} = \frac{75 + 78}{2} = 76{,}5 \text{ млрд руб.},$$

где 78 млрд руб. — стоимость основных средств на конец года (78 = 75 + 2 − 7 + 8).

Таблица 8.5

Даты времени	Стоимость основных средств (в млрд руб.)	Число месяцев, в течение которых стоимость не изменялась	$y_i t_i$
I/I	75	3	225
I/IV	77	2	154
I/VI	70	4	280
I/X	78	3	234
Итого		12	893

Вот почему для анализа деятельности предприятия следует пользоваться данными внутренней отчетности, что позволяет получить более достоверную оценку результатов.

При определении средних уровней временного ряда нужно иметь в виду, что средняя будет достаточно надежной характеристикой ряда динамики, если она характеризует период с более или менее стабильными условиями развития. Если же за исследуемый период можно выделить этапы, в течение которых условия развития существенно менялись, то пользоваться общей средней не всегда целесообразно, а предпочтение нужно отдать средним, рассчитанным по отдельным подпериодам.

Средний абсолютный прирост (или средняя скорость роста) рассчитывается как средняя арифметическая из показателей скорости роста за отдельные промежутки времени:

$$\Delta = \frac{\sum_{i=1}^{n-1} \Delta_i}{n-1}, \qquad (8.9)$$

где n — число уровней ряда;

Δ_i — абсолютные изменения по сравнению с предшествующим уровнем. По данным табл. 8.4 $\bar{\Delta} = \frac{-1,1}{4} = -0,275$ млн кв. м.

Это означает, что в течение 1990—1994 гг. в среднем объем строительства жилья жилищно-строительными кооперативами в России снижался на 275 млн кв. м.

Формулу (8.9) можно преобразовать:

$$\sum_{i=1}^{n-1} \Delta_i = y_n - y_1,$$

$$\text{тогда } \bar{\Delta} = \frac{y_n - y_1}{n-1}, \qquad (8.9а)$$

где y_n и y_1 — соответственно конечный и начальный уровни динамического ряда.

Средний коэффициент роста вычисляется по формуле средней геометрической из показателей коэффициентов роста за отдельные периоды:

$$\bar{K} = \sqrt[n-1]{K_1 \times K_2 \times \ldots \times K_{n-1}}, \qquad (8.10)$$

где K_1, K_2, ..., K_{n-1} — коэффициенты роста по сравнению с уровнем предшествующего периода;

n — число уровней ряда.

Эту формулу можно записать иначе, если использовать взаимосвязь коэффициентов роста, вычисленных с переменной базой, и коэффициентов роста, рассчитанных с постоянной базой, т.е. учитывая, что $K_1 \times K_2 \times K_3 \times \ldots \times K_{n-1} = \frac{y_n}{y_1}$, средний коэффициент роста можно определить по формуле:

$$\bar{K} = \sqrt[n-1]{\frac{y_n}{y_1}}. \qquad (8.10а)$$

По данным табл. 8.4, средний коэффициент роста будет равен:

$$\bar{K} = \sqrt[3]{\frac{0,6207}{0,8276}} = 0,91,$$

т.е. формула (8.10а) позволяет определить средний коэффициент роста на основе данных о начальном и конечном уровнях ряда динамики. Средний коэффициент роста показывает, во сколько раз в среднем за отдельные составляющие рассматриваемого периода изменялись уровни динамического ряда.

Средний темп роста представляет собой средний коэффициент роста, выраженный в процентах:

$$\bar{T} = \bar{K} \times 100\%,$$

где $\bar{K}$ — средний годовой коэффициент роста.

Для приводимого примера $\bar{T} = 88,76\%$, а это означает, что в среднем ежегодно строительство жилья составляло 88,76% к уровню предыдущего года. Более доходчива интерпретация относительных изменений с помощью среднего темпа прироста $(\bar{T} = \bar{T} - 100\%)$, т.е., продолжая пример, можно сказать, что в среднем ежегодно строительство жилья жилищно-строительными кооперативами в России за 1990—1994 гг. снижалось на 11,24%.

Для практического применения средний темп роста, рассчитанный по данным о конечном и начальном уровнях временного ряда, можно использовать только в случае более или менее равномерного изменения уровней. В случае сильной колеблемости уровней ряда использование средней геометрической может дать искаженное выражение средней интенсивности изменения их уровней. В таких случаях предлагается средний коэффициент роста рассчитывать как параболический средний темп роста, средний геометрический темп роста выравненного ряда показательной функции и т.д.

Особую осторожность при применении средних абсолютных приростов или средних темпов роста (прироста) следует соблюдать в тех случаях, когда появляется перелом в имевшей место тенденции изменения уровней динамического ряда. Например, известны данные о производстве приборов по месяцам полугодия (табл. 8.6).

Таблица 8.6

	Январь	Февраль	Март	Апрель	Май	Июнь
Произведено приборов, (в тыс. шт.)	2,0	2,1	2,3	2,6	2,8	1,9
Темпы роста (в % к предыдущему месяцу)	100,0	105,0	109,5	113,0	107,7	67,9

Средний месячный коэффициент роста объема выпуска за первое полугодие составит 92,54% $\left(\sqrt[5]{0{,}679} \times 100\%\right)$, т.е. по результатам расчетов в среднем за полугодие ежемесячно выпуск продукции снижался на 7,46%. Этот результат противоречит приведенным темпам роста объема выпуска за январь—май отчетного года. В действительности за январь—май ежемесячно объем продукции увеличивался в среднем на 8,8% $\left(\sqrt[4]{\frac{2{,}8}{2{,}0}} \times 100\% - 100\%\right)$, что наглядно видно на рис. 8.2.

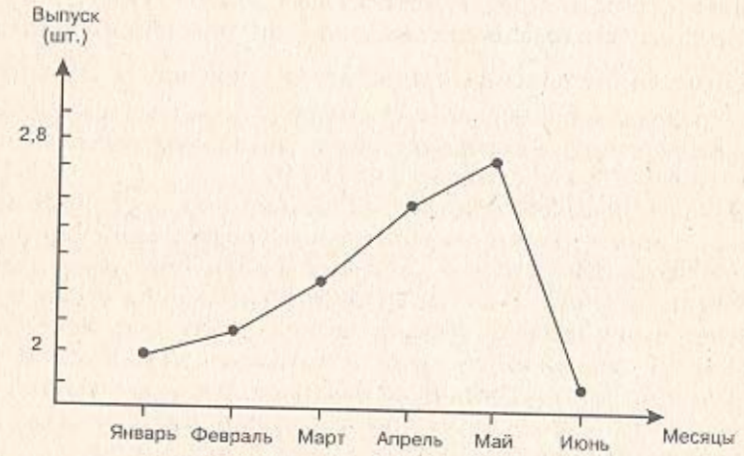


Рис. 8.2. Производство приборов за первое полугодие

Резкий спад объема производства приборов в июне может означать либо резкое изменение общей тенденции, либо влияние каких-то случайных факторов. Поэтому в данном случае можно ориентироваться на средний темп роста за период с января по май, когда средний месячный темп роста составит 108,8% $\left(\sqrt[4]{\frac{2{,}8}{2{,}0}} \times 100\%\right)$.

Для последующих периодов следует выполнить самостоятельный анализ.

8.4. Каковы характеристика основной тенденции развития и способы ее выявления?

Одной из задач, возникающих при анализе рядов динамики, является установление закономерности изменения уровней изучаемого показателя во времени.

Уровни ряда динамики формируются под совокупным влиянием множества длительно и кратковременно действующих факторов и в том числе различного рода случайных обстоятельств. Выявление основной закономерности изменения уровней ряда предполагает ее количественное выражение, в некоторой мере свободное от случайных воздействий. *Выявление основной тенденции развития (тренда) называется в статистике также выравниванием временного ряда, а методы выявления основной тенденции — методами выравнивания.* Выравнивание позволяет характеризовать особенность изменения во времени данного динамического ряда в наиболее общем виде как функцию времени, предполагая, что через время можно выразить влияние всех основных факторов.

Наглядно представить процесс развития явления во времени позволяет графическое изображение уровней временного ряда. Анализ общей тенденции развития явления чаще всего производится с использованием линейных диаграмм. Для их построения на оси абсцисс располагают отрезки, соответствующие определенным моментам или периодам времени, а на оси ординат — уровни ряда динамики. Рассмотрим на рис. 8.3 графическое представление данных об урожайности зерновых культур в Российской Федерации за 1970—1998 гг. (данные приведены в табл. 8.7).

Уже по виду графика можно предполагать наличие разных тенденций в динамике урожайности на протяжении разных временных периодов. Визуально можно выделить три подпериода: первый — с 1970 по 1980 г.; второй — с 1981 по 1989 г. и третий — с 1990 по 1998 г.

Таблица 8.7

Урожайность зерновых культур в Российской Федерации (в центнерах с гектара)

Год	Урожайность	Год	Урожайность
1970	14,8	1985	14,5
1971	13,8	1986	15,9
1972	11,8	1987	14,8
1973	15,8	1988	14,2
1974	13,7	1989	16,4
1975	9,4	1990	18,5
1976	15,4	1991	14,4
1977	13,0	1992	17,2
1978	16,5	1993	16,3
1979	11,2	1994	14,4
1980	12,9	1995	11,6
1981	10,0	1996	12,9
1982	13,6	1997	16,5
1983	14,8	1998	9,4
1984	12,2		

Источник: Вопросы статистики. 1999. № 9. С. 83.

Первый подпериод характеризуется достаточно высокой колеблемостью уровней относительно средней и слабым влиянием фактора времени на изменение уровня урожайности. Во втором периоде наблюдается в целом тенденция роста урожайности с течением времени. А вот в третий период можно говорить об общей тенденции снижения урожайности зерновых.

Задача аналитика в данном случае состоит, с одной стороны, в измерении колеблемости уровней, а с другой — в моделировании тенденции изменения этих уровней во времени.

Представив на графике линейные уравнения тренда[24], рассчитанные для каждого подпериода, мы получим наглядное отображение тенденции изменения урожайности в каждом временном интервале. На уровень ряда динамики, помимо систематических и случайных факторов, влияют также причины, обусловленные периодичностью колебаний.

[24] Для 1970—1980 гг. уравнение тренда имеет вид $\tilde{y}_t = 0{,}14 - 0{,}0882 \cdot t$ ($t = 1, 2,3,...,11$). Для временного интервала 1981—1989 гг. $\tilde{y}_t = 6{,}0 + 0{,}4983 \cdot t$ ($t = 12, 13,...,20$). И наконец, тенденция за 1989—1998 гг. может быть представлена уравнением $\tilde{y}_t = 37{,}2 - 0{,}7233 \cdot t$ ($t = 21, 22,...,29$). (См. Манелля А.И. Вопросы статистики. 1999. № 9. С. 83.)



Рис. 8.3. Урожайность зерновых культур в Российской Федерации

Примечание. y_t — расчетная урожайность в t-м году, определенная по уравнению локального тренда.

В рядах внутригодовой динамики можно выделить три важнейшие составляющие колеблемости уровней временного ряда: тренд, сезонный и случайный компоненты. На рис. 8.4 показаны вышеназванные виды колебаний.

При анализе деловой активности на протяжении значительного числа периодов можно выделить еще один вид — циклические колебания, отличающиеся от сезонных отсутствием регулярной модели динамики (рис. 8.5).

На основе анализа фактических данных можно констатировать существование циклов процентных ставок, однако трудно предсказать как пик, так и время, и направление изменений процентных ставок. Например, наметившаяся тенденция снижения британской базовой процентной ставки в 1974 г. была прервана резким ее ростом в 1977 г., сменившимся падением до 6,5% в 1978 г.

Таким образом, при анализе колеблемости динамических рядов наряду с выделением случайных колебаний возникает и задача изучения периодических колебаний. Как правило, изучение периоди-

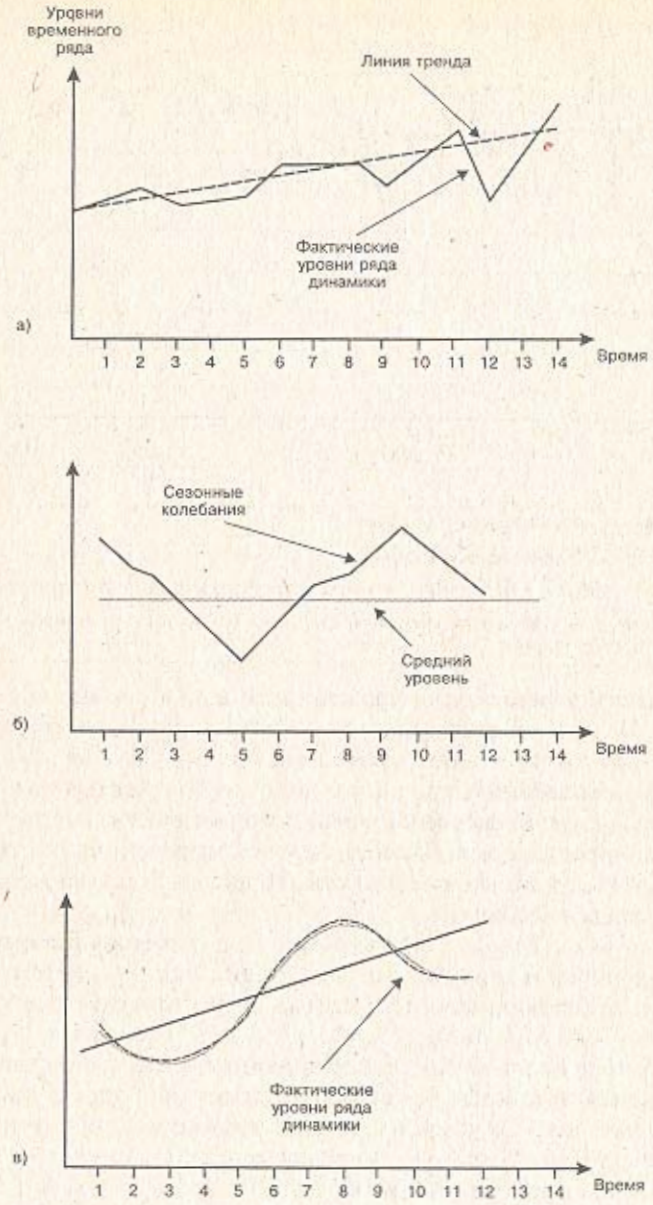


Рис. 8.4. Виды колебаний уровней временного ряда



Рис. 8.5. Британская базовая ставка

Источник: Райе Т., Койли Б. Финансовые инвестиции и риск. Кельн: Торгово-издательское бюро BHV, 1995. С. 394.

ческих (сезонных) колебаний необходимо с целью исключения их влияния на общую динамику для выявления чистой (случайной) колеблемости.

8.5. Какова роль аналитического выравнивания уровней временного ряда?

Для того чтобы представить количественную модель, выражающую общую тенденцию изменений уровней динамического ряда во времени, используется ***аналитическое выравнивание ряда динамики***. В этом случае фактические уровни заменяются уровнями, вычисленными на основе определенной кривой. Предполагается, что она отражает общую тенденцию изменения во времени изучаемого показателя.

При аналитическом выравнивании ряда динамики закономерно изменяющийся уровень изучаемого показателя оценивается как функция времени $\tilde{y}_t = f(t)$, где $\tilde{y}_t$ — уровни динамического ряда, вычисленные по соответствующему аналитическому уравнению на момент времени t. В табл. 8.8 приводятся различные виды трендовых моделей, наиболее часто используемые для аналитического выравнивания.

Выбор формы кривой во многом определяет результаты экстраполяции тренда. Основанием для выбора вида кривой может быть содержательный анализ сущности развития данного явления. Можно опираться на результаты предыдущих исследований в данной области, а также на анализ графического изображения уровней динамического ряда (линейной диаграммы). Однако из графического представления эмпирических данных не всегда удается произвести однозначный выбор формы уравнения. Поэтому целесообразно воспользоваться графическим изображением сглаженных уровней, в которых случайные и волнообразные колебания в некоторой степени оказываются погашенными.

При выборе вида кривой для выравнивания динамического ряда возможно также использование ***метода конечных разностей***, который основан на свойствах различных кривых, применяемых при выравнивании.

Одним из наиболее распространенных приемов сглаживания уровней первоначального ряда динамики является ***метод скользящей средней***. Для определения скользящей средней формируем укрупненные интервалы, состоящие из одинакового числа уровней. Каждый последующий интервал получаем, постепенно сдвигаясь от начального уровня динамического ряда на один уровень. Тогда первый интервал будет включать уровни $y_1, y_2, \ldots y_m$; второй — уровни $y_2, y_3 \ldots y_{m+1}$ и т.д. Таким образом, интервал сглаживания как бы скользит по динамическому ряду с шагом, равным единице. По сформированным укрупненным интервалам определяем сумму значений уровней, на основе которых рассчитываем скользящие средние. Полученная средняя относится к середине укрупненного интервала. Поэтому при сглаживании скользящей средней технически удобнее укрупненный интервал составлять из нечетного числа уровней ряда. Нахождение скользящей средней по четному числу уровней создает неудобство, вызываемое тем, что средняя может быть отнесена только к середине между двумя датами. В этом случае необходима дополнительная процедура центрирования средних. Как следует подходить к выбору интервала сглаживания? Нередко выбор интервала сглаживания осуществляется произвольно, однако при этом нужно учитывать количество уровней в анализируемом ряду динамики, так как при использовании приема скользящей средней сглаженный ряд сокращается по сравнению с исходным рядом на число уровней, равное $(m-1)$. Вместе с тем, чем продолжительнее интервал сглаживания, тем сильнее усреднение, поэтому выявляемая тенденция развития получается более плавной. Чаще всего интервал сглаживания может состоять из трех, пяти или семи уровней.

Таблица 8.8

Название функции	Описание функции
1. Линейная	$\hat{y}_t = b_0 + b_1 t$
2. Парабола второго порядка	$\hat{y}_t = b_0 + b_1 t + b_2 t^2$
3. Кубическая парабола	$\hat{y}_t = b_0 + b_1 t + b_2 t^2 + b_3 t^3$
4. Показательная	$\hat{y}_t = b_0 b_1^t$
5. Экспоненциальная	$\hat{y}_t = b_0 e^{b_1 t}$
6. Модифицированная экспонента	$\hat{y}_t = b_0 + b_1 b_2^t$
7. Кривая Гомперца	$\hat{y}_t = b_0 b^{b_2^t}$
8. Логистическая кривая	$\hat{y}_t = \frac{b_0}{1 + b_1 e^{-b_2 t}}$
9. Логарифмическая парабола	$\hat{y}_t = b_0 b_1^t b_2^{t^2}$
10. Гиперболическая	$\hat{y}_t = b_0 + b_1 \frac{1}{t}$

Основываясь на свойствах конечных разностей для различных видов кривых, Б.С. Ястремский[25] сделал вывод о применимости для выравнивания линейной функции, если любые три равностоящие уровня имеют нулевую вторую разность. Порядок разностей, остающихся примерно равными друг другу, принимается за степень выравнивающего многочлена, т.е. если примерно одну и ту же величину имеют вторые разности, то для выравнивания используется парабола второго порядка.

Можно указать и ряд других признаков, которые могут помочь при выборе формы кривой: если примерно постоянными оказываются темпы роста, то для выравнивания применяется показательная функция; если первые разности имеют тенденцию уменьшаться с постоянным темпом, то следует остановиться на модифицированной экспоненте; если средние уровни, нанесенные на полулогарифмическую сетку, близки к прямой линии, то предпочтительнее простая экспонента; если первые разности обратных зна-

[25] Ястремский Б.С. Некоторые вопросы математической статистики. 3-е изд. М.: Госстатиздат, 1961.

чений средних уровней изменяются на один и тот же процент, то следует остановиться на логистической кривой.

При выборе формы уравнения следует исходить и из объема имеющейся информации. Чем больше параметров содержит уравнение тренда, тем больше должно быть наблюдений при одной и той же степени надежности оценивания.

Выбор формы кривой может осуществляться и на основе принятого критерия, в качестве которого может служить сумма квадратов отклонений фактических значений от значений, рассчитанных по уравнению тренда. Из совокупности кривых выбирается та, которой соответствует минимальное значение критерия.

Рассмотрим аналитическое выравнивание ряда динамики по прямой, т.е. аналитическое уравнение вида:

$$\bar{y}_t = b_0 + b_1 t,$$

где t — порядковый номер периодов или моментов времени.

Параметры b_0 и b_1 прямой рассчитываются по ***методу наименьших квадратов*** (МНК). Система нормальных уравнений в данном случае имеет вид:

$$\begin{cases} \sum_{i=1}^{n} y_i = b_0 n + b_1 \sum_{i=1}^{n} t_i \\ \sum_{i=1}^{n} y_i t_i = b_0 \sum_{i=1}^{n} t_i + b_1 \sum_{i=1}^{n} t_i t_2^i \end{cases}.$$

Расчет параметров уравнения можно упростить, если отсчет времени производить так, чтобы сумма показателей времени изучаемого ряда динамики была равна нулю $\left(\sum_{i=1}^{n} t_i = 0\right)$. При нечетном числе уровней ряда динамики для получения $\sum_{i=1}^{n} t_i = 0$ уровень, находящийся в середине ряда, принимается за условное начало отсчета времени (этому периоду или моменту времени придается нулевое значение). Даты времени, стоящие выше этого уровня, обозначаются натуральными числами со знаком минус (−1, −2, −3 и т.д.), а ниже — натуральными числами со знаком плюс (+1, +2, +3 и т.д.).

Если число уровней динамического ряда четное, периоды времени верхней половины ряда (до середины) нумеруются −1, −3, −2 и т.д., а нижней — +1, +3, +5 и т.д. При этом условии $\sum_{i=1}^{n} t_i$ будет равна нулю, и система нормальных уравнений преобразуется следующим образом:

$$\begin{cases} \sum_{i=1}^{n} y_i = b_0 n \\ \sum_{i=1}^{n} y_i t_i = b_1 \sum_{i=1}^{n} t_i t_2^i \end{cases},$$

откуда $b_0 = \dfrac{\sum_{1}^{n} y_i}{n}$; $b_1 = \dfrac{\sum_{1}^{n} y_i t_i}{\sum_{1}^{n} t_i^2}$.

Рассмотрим аналитическое выравнивание по прямой ряда динамики строительства жилья жилищно-строительными кооперативами в России (см. данные табл. 8.4). Расчет параметров уравнения прямой представлен в табл. 8.9.

Используя итоги граф 2, 4 и 5, определим параметры уравнения прямой:

$$b_0 = \frac{11{,}1}{5} = 2{,}22; \; b_1 = \frac{-2{,}7}{10} = -0{,}27.$$

Таблица 8.9

Годы	Построено жилищно-строительными кооперативами (в млн м²) y_i	Условные обозначения периодов t_i	$y_i t_i$	t_i^2	Выравненные уровни ряда динамики (в млн м²)	$y_i - \bar{y}_i$	$(y_i - \bar{y}_i)^2$
1	2	3–2	4	5	6	7	8
1990	2,9	−1	−5,8	4	2,76	0,14	0,0196
1991	2,4		−2,4	1	2,49	−0,09	0,0081
1992	2,1	0	0	0	2,22	−0,12	0,0144
1993	1,9	+1	1,9	1	1,95	−0,05	0,0025
1994	1,8	+2	3,6	4	1,68	0,12	0,0144
Итого	11,1		−2,7	10	11,10	0,00	0,0590

По рассчитанным параметрам записывается уравнение прямой ряда динамики, характеризующее строительство жилья жилищно-строительными кооперативами:

$$\bar{y}_t = 2{,}22 - 0{,}27t .$$

Используя приведенное уравнение, рассчитаем для каждого года теоретические значения:

для 1990 г. $\bar{y}_{t=-2} = 2{,}22 - 0{,}27 \times (-2) = 2{,}76$ млн м²;

для 1991 г. $\bar{y}_{t=-2} = 2{,}22 - 0{,}27 \times (-1) = 2{,}49$ млн м² и т.д. (см. графу 6 табл. 8.9).

Правильность расчета уровней выравниваемого ряда динамики может быть проверена следующим образом: сумма значений эмпирического ряда должна совпадать с суммой вычисленных уровней выравненного ряда, т.е. $\sum_{i=1}^{n} \bar{y}_i = \sum_{i=1}^{n} \bar{y}_t$ (см. итоги граф 2 и 6).

8.6. Как выполнить прогноз по уравнению тренда?

Продление в будущее тенденции, наблюдавшейся в прошлом, носит название ***экстраполяции***. Экстраполируя при $t = 3$, находим уровень 1995 г., равный 1,41 млн м² (2,22 − 0,27×3).

Возможность экстраполяции обеспечивается двумя обстоятельствами:

1) общие условия, определяющие тенденцию развития в прошлом, не претерпевают существенных изменений в будущем;

2) тенденция развития явления характеризуется тем или иным аналитическим уравнением.

Общая тенденция развития может быть охарактеризована с помощью содержательного экономического анализа. Вместе с тем расчет таких показателей, как скорость роста, темпы роста, пункты роста, позволяет определить наличие или отсутствие устойчивой тенденции развития и обосновать форму уравнения тренда. Если условия формирования уровней ряда изменяются, то расчет параметров уравнения не следует вести по данным за весь рассматриваемый период времени. В этом случае целесообразно разбить ряд динамики на этапы, ориентируясь на устойчивость абсолютных приростов или пунктов роста. Значение $\bar{y}_t$, полученное в результате экстраполяции, используют для определения прогнозного значения на будущее.

При составлении прогнозов оперируют не точечной, а интервальной оценкой, определяя так называемые ***доверительные интервалы прогноза***. Величина доверительного интервала определяется в общем виде так:

$$\bar{y}_t = \pm t_\alpha \frac{S_{\bar{y}}}{\sqrt{n}}, \qquad (8.13)$$

где $S_{\bar{y}}$ — среднее квадратическое отклонение от тренда;

t_α — табличное значение r-критерия Стьюдента при уровне значимости α.

Величина $S_{\bar{y}}$ определяется по формуле:

$$S_{\bar{y}} = \sqrt{\frac{\sum_{1}^{n} (y_i - \bar{y}_t)^2}{n - m}}, \qquad (8.14)$$

где y_i и $\bar{y}_t$ — соответственно фактическое и расчетное значения уровней динамического ряда;

n — число уровней ряда;

m — количество параметров в уравнении тренда (для уравнения прямой $m = 2$).

Используя данные графы 8 табл. 8.9, рассчитаем среднюю квадратическую ошибку линейного уравнения тренда:

$$S_{\bar{y}} = \sqrt{\frac{0{,}0590}{5-2}} = 0{,}14 \text{ млн м}^2.$$

Отсюда величина относительной ошибки составляет 6,32%:

$$\left(\frac{S_{\bar{y}}}{\bar{y}} \cdot 100\% = \frac{0{,}14}{2{,}22} \cdot 100\% \right).$$

Величина средней квадратической ошибки может быть рассчитана иным способом, позволяющим избежать ошибки при округлениях величины $(y_i - \bar{y}_t)$. Для уравнения линейного тренда $\bar{y}_t = b_0 + b_1 \delta y$ величина $S_{\bar{y}}$ определяется по формуле:

$$S_{\bar{y}} = \sqrt{\sum_{i=1}^{n} y_i^2 - b_0 \sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} y_i t} . \qquad (8.14а)$$

Для примера, приведенного в табл. 8.9, получим $S_y = 0{,}14$ млн м².

$$S_y = \sqrt{\frac{25{,}43 - 2{,}22 \cdot 11{,}1 - 0{,}27 \cdot 2{,}7}{n-2}} = 0{,}14 \text{ млн м}^2.$$

Если воспользоваться методом конечных разностей для выбора формы уравнения тренда, то из данных графы 4 табл. 8.5 можно видеть, что абсолютные приросты не являются постоянными, а начиная с 1992 г., постоянными будут вторые разности (разности абсолютных приростов). В этом случае, как было показано выше, для выравнивания может использоваться парабола второго порядка: $\bar{y}_t = b_0 + b_1 t + b_2 t^2$.

Система нормальных уравнений для нахождения параметров уравнения параболы (при соблюдении принципа отсчета от условного начала) будет иметь вид:

$$\begin{cases} \sum\limits_{i=1}^{n} y_i = b_0 n + b_2 \sum\limits_{1}^{n} t_i^2 \\ \sum\limits_{i=1}^{n} y_i t_i = b_1 \sum\limits_{1}^{n} t_i t_2^i \\ \sum\limits_{i=1}^{n} y_i t_i^2 = b_0 \sum\limits_{1}^{n} t_1^2 + b_2 \sum\limits_{1}^{n} t_i^4 \end{cases}.$$

Расчет параметров этого уравнения тренда представлен в табл. 8.10.

Подставляем итоги графы 2, 4, 5, 6 и 7 табл. 8.10 и получаем следующую систему уравнений для данного временного ряда:

Таблица 8.10

Годы	y_i	t_i	t_i^2	$y_i t_i$	$y_i t_i^2$	t_i^4	$\bar{y}_i$	$y_i - \bar{y}_i$
1	2	3	4	5	6	7	8	9
1990	2,9	−2	4	−5,8	11,6	16	2,889	+0,011
1991	2,4	−1	1	−2,4	2,4	1	2,426	−0,026
1992	2,1	0	0	0	0	0	2,091	+0,009
1993	1,9	+1	1	1,9	1,9	1	1,886	+0,014
1994	1,8	+2	4	3,6	7,2	16	1,809	−0,009
Итого	11,1	0	10	-2,7	23,1	34	11,10	0,00

$$\begin{cases} 5b_0 - 10b_2 = 11{,}1 \\ 10b_1 = -2{,}7 \\ 10b_0 + 34b_2 = 23{,}1 \end{cases}$$

Решая систему уравнений, определим значения параметров уравнения параболы второго порядка:

$b_0 = 2{,}0914$; $b_1 = -0{,}27$; $b_2 = 0{,}0643$.

Отсюда уравнение параболы второго порядка, характеризующее тенденцию строительства жилья жилищно-строительными кооперативами, будет записано так:

$$\bar{y}_t = 2{,}0914 - 0{,}27t + 0{,}0643t^2.$$

Величина среднего квадратического отклонения фактических уровней динамического ряда от выравненных для уравнения параболы второго порядка определяется по формуле:

$$S_y = \sqrt{\frac{\sum\limits_{i=1}^{n} y_i^2 - b_0 \sum\limits_{i=1}^{n} y_i - b_1 \sum\limits_{i=1}^{n} y_i t_i - b_2 \sum\limits_{i=1}^{n} y_i t^2}{n-3}},$$

т.е. $S_y = \sqrt{\frac{25{,}4 - 2{,}0914 \cdot 11{,}1 - 0{,}27 \cdot 2{,}7 - 0{,}0643 \cdot 23{,}1}{5-3}} =$

$= \sqrt{0{,}00057} = 0{,}02377$ млн м².

Относительная ошибка уравнения составит 1,07%. Сравнив полученные значения S_y для уравнения прямой и параболы второго порядка, можно сделать вывод о том, что парабола более точно описывает основную тенденцию ряда динамики, характеризующего строительство жилья жилищно-строительными кооперативами.

Прогноз на 1995 г. по уравнению параболы второго порядка будет составлен следующим образом:

1) экстраполируя для $t = 3$, получаем $\bar{y}_t = 1{,}860$ млн м²;

2) значение критерия Стьюдента при уровне значимости 5% и числе степеней свободы $n - m = 2$ равно 4,303 (см. Приложение 4);

3) $t_\alpha \times \frac{S_y}{\sqrt{n}} = 4{,}303 \frac{0{,}02377}{\sqrt{5}} = 0{,}04574$ млн м²;

4) $1{,}860 - 0{,}046 \le y_{\text{прогноз}} \le 1{,}860 + 0{,}046$;

$1{,}814 \le y_{\text{прогноз}} \le 1{,}906$.

Таким образом, с вероятностью 95% можно ожидать, что в 1995 г. жилищно-строительные кооперативы должны были построить жилья не меньше, чем 1,814 млн м², но не больше 1,906 млн м².

Приведенные расчеты следует рассматривать не как завершающую стадию прогнозирования, а лишь как предварительный этап в разработке прогноза. Для составления прогноза должна быть привлечена дополнительная информация, не содержащаяся в самом динамическом ряду.

Если в изменениях уровней обнаруживается тенденция к постоянству темпов роста, то выравнивание ряда следует проводить по показательной кривой:

$$\bar{y}_t = b_0 b_1^t,$$

где b_0 — коэффициент роста.

Техника выравнивания по показательной кривой аналогична технике выравнивания по прямой, за исключением того, что выравниваются не уровни ряда, а их логарифмы:

$$\begin{cases} \sum_{i=1}^{n} \lg y = n \lg b_0 + \lg b_1 \sum_{1}^{n} t \\ \sum_{i=1}^{n} \lg y \cdot t = \lg b_0 \sum_{1}^{n} t + \lg b_1 \sum_{1}^{n} t^2 . \end{cases}$$

Приравнивая $\sum t$ к нулю, получаем:

$$\begin{cases} \sum_{i=1}^{n} \lg y = n \lg b_0 \\ \sum_{i=1}^{n} \lg y \cdot t = \lg b_1 \sum_{i=1}^{n} t^2 . \end{cases}$$

По вычисленным значениям логарифмов определяем величины параметров уравнения показательной кривой b_0 и b_1.

Современные компьютерные программы по анализу временных рядов позволяют автоматически определять тип модели, адекватной исходным данным, на основе соответствующего критерия.

Аналитическое уравнение представляет собой математическую модель развития явления и отражает статистическую закономерность, проявляющуюся в рядах динамики. Следует помнить, что прием аналитического выравнивания содержит некоторые условности, связанные прежде всего с тем, что уровни, характеризующие тот или иной динамический ряд, рассматриваются как функция времени. В действительности же развитие явлений обусловлено не тем, сколько времени прошло с отправного момента, а тем,

какие факторы влияли на развитие, в каком направлении и с какой интенсивностью. Развитие явлений во времени выступает как внешнее выражение этих факторов, как их суммарное действие, оказывающее влияние на изменение уровня в отдельно взятые промежутки или моменты времени. Выявить основную тенденцию развития явления методом наименьших квадратов можно лишь тогда, когда выяснено, что изменяющиеся во времени процессы протекают в течение всего рассматриваемого промежутка времени одинаково, что их количественное и качественное изменение происходит под действием одного и того же комплекса основных факторов, определяющих движение данного ряда динамики.

Использование экстраполяции в изменившихся условиях, появление новых факторов, оказывающих влияние на формирование уровней временного ряда, будет сопряжено с более или менее значительными ошибками. Причем экстраполяция на отдаленные даты подвержена более значительным ошибкам, чем краткосрочная экстраполяция. Это в известной степени может объясняться определенной инерционностью развития управляемых объектов: чем выше уровень управления, тем более инерционен объект управления. Однако с удалением периода прогноза от фактических уровней временного ряда влияние инерционности развития снижается и возрастает влияние новых условий и факторов. Поэтому целесообразно постоянно обновлять временные ряды по мере получения новых данных, что обусловливает и корректировку уравнений тренда.

Выделим особенности моделей аналитического выравнивания уровней динамического ряда, накладывающие ограничения на их использование. Во-первых, динамические ряды, к которым применяется аппроксимация, должны быть достаточно длинными. Во-вторых, применение аппроксимации наиболее целесообразно в случае медленно и плавно меняющегося уровня. В-третьих, аппроксимация как метод моделирования практически не адаптируется к изменяющимся условиям формирования уровней ряда; при появлении новых данных построение модели должно быть проведено заново. В-четвертых, при использовании для расчета параметров уравнения метода наименьших квадратов (МНК) считается, что значимость информации в пределах отрезка аппроксимации одинакова, независимо от давности полученных данных, в то время как более поздние данные имеют большую ценность.

Помимо этого, динамические ряды экономических показателей часто имеют небольшую длину и подвержены значительным колебаниям, которые аппроксимация предвидеть не может. Поэтому в практике статистического анализа экономических процессов большое распространение получили ***методы адаптивного моделирования и прогнозирования***.

8.7. Что такое сезонная неравномерность и какова ее характеристика?

В широком понимании к ***сезонным*** относят все явления, которые обнаруживают в своем развитии отчетливо выраженную закономерность внутригодовых изменений, т.е. более или менее устойчиво повторяющиеся из года в год колебания уровней. Часто эти колебания не обязательно связаны со сменой времен года. К сезонным явлениям относят, например, потребление электроэнергии; неравномерность производственной деятельности в отраслях пищевой промышленности, связанных с переработкой сельскохозяйственного сырья; перевозки пассажирским транспортом; спрос на многие виды продукции и услуг и т.д.

Как бы ни проявлялась сезонность, она наносит определенный ущерб национальной экономике, связанный с неравномерным использованием оборудования и рабочей силы, неравномерной загрузкой транспорта, необходимостью создания резервов мощностей и т.д. Комплексное регулирование сезонных изменений по отдельным отраслям должно основываться на исследовании сезонных отклонений.

Важнейшими задачами, решаемыми в ходе исследования сезонности, являются следующие:

- определение наличия сезонности, численное выражение проявления сезонных колебаний и выявление их силы и характера в различных фазах годового цикла;
- характеристика факторов, вызывающих сезонные колебания;
- оценка последствий, к которым приводит наличие сезонных колебаний;
- математическое моделирование сезонности.

Для измерения ***сезонных колебаний*** статистикой предложены различные методы. Наиболее простые и часто употребляемые из них:

- метод абсолютных разностей;
- метод относительных разностей;
- построение индексов сезонности.

Первые два способа предполагают нахождение разностей фактических уровней и уровней, найденных при выявлении основной тенденции развития.

Применяя способ абсолютных разностей, оперируют непосредственно размерами этих разностей, а при использовании метода относительных разностей определяют отношение абсолютных размеров указанных разностей к выравненному уровню. При выявлении основной тенденции используют либо метод скользящей средней, либо аналитическое выравнивание. В некоторых случаях в стационарных рядах можно пользоваться разностью фактических уровней и средним месячным уровнем за год. Для измерения сезонных колебаний необходимы данные не менее чем за трехлетний период. Использование данных за несколько лет связано с тем обстоятельством, что в отклонениях по отдельным годам сезонные колебания смешиваются со случайными. Чтобы элиминировать случайные колебания, берут средние отклонения за несколько лет.

Важной характеристикой сезонных колебаний является индекс сезонности. В простейшем варианте этот показатель рассчитывается как отношение среднего уровня соответствующего месяца к общей средней за все месяцы нескольких лет.

Индексы сезонности могут быть рассчитаны также как отношение фактического уровня соответствующего месяца к уровню, рассчитанному по методу скользящей средней или же определенному по уравнению тренда. В последнем случае предварительно для временного ряда рассчитывается уравнение тренда, на основании которого за каждый месяц определяется значение выравненного уровня.

Тогда индекс сезонности рассчитывается следующим образом:

$$i_{сез} = \frac{y_i}{\hat{y}_i} \cdot 100\%, \qquad (8.19)$$

где $\hat{y}_i$ — значение выравненного уровня i-го месяца.

Рассмотрим выявление всех типов колебаний внутригодовой динамики уровней временного ряда на следующем примере, характеризующем динамику квартального товарооборота специализированного магазина зарубежной фирмы за пять лет (табл. 8.11).

Для выравнивания уровней ряда динамики товарооборота принимаем период сглаживания, равный четырем кварталам ($m = 4$).

Первая скользящая средняя равна 16 тыс. долл. $\left(\frac{16+21+9+18}{4}\right)$, но отнесена она будет не к конкретному кварталу, а попадет в промежуток между вторым и третьим кварталом года. Точно так же к промежутку между кварталами отнесены и все последующие скользящие средние (см. графу 4 табл. 8.11). Для отнесения скользящей средней к определенному кварталу находим средние из двух смежных скользящих средних, т.е. производим центрирование средних. Так, для третьего квартала 1994 г. центрированная скользящая сред-

Таблица 8.11

Год	Квартал	Фактический объем товарооборота (в тыс. долл.)	Скользящая средняя за четыре квартала (в тыс. долл.)	Центрированная скользящая средняя (в тыс. долл.)	Отношение фактического объема товарооборота к скользящей средней (гр.3: гр.5)	Индекс сезонности	Объем товарооборота, скорректированный на индекс сезонности (в тыс. долл.) (гр.3: гр.7)
1	2	3	4	5	6	7	8
1994	I	16		—		0,9506	16,8
	II	21	16,00		—	1,2988	16,2
	III	9	15,75	15,875	0,567	0,6119	14,7
	IV	18	15,70	15,625	1,152	1,1387	15,8
1995	I	15	15,75	15,625	0,960	0,9506	15,8
	II	20	5,75	15,750	1,270	0,2988	15,4
	III	10	16,25	16,000	0,625	0,6119	16,3
	IV	18	17,25	16,750	1,075	1,1387	15,8
1996	I	17	18,00	17,625	0,965	0,9506	17,9
	II	24	19,00	18,500	1,297	1,2988	18,5
	III	13	19,00	19,000	0,684	0,6119	21,2
	IV	22	19,25	19,125	1,150	1,1387	19,3
1997	I	17	18,75	19,000	0,895	0,9506	17,9
	II	25	18,50	18,625	1,342	1,2988	19,2
	III	11	18,75	18,625	0,591	0,6119	18,0
	IV	21	19,00	18,875	1,113	1,1387	18,4
1998	I	18	19,75	19,375	0,929	0,9506	18,9
	II	26	20,75	20,250	1,284	1,2988	20,0
	III	14		—		0,6119	22,9
	IV	25		—		1,1387	21,9

няя будет равна 15,875 тыс. долл. $\left(\frac{16,00 + 15,75}{2}\right)$. Аналогичным образом определяются и все остальные средние, записанные в графе 5 табл. 8.11.

Для выявления сезонной составляющей в колеблемости уровней ряда динамики рассчитываем отношения фактических объемов товарооборота каждого квартала к соответствующей скользящей средней. Эти соотношения представлены в графе 6 табл. 8.11 [графа 6 = графа 3 : графа 5]. Для примера находим относительный показатель для III квартала 1994 г.:

Фактический объем товарооборота/Центрированная скользящая средняя = 9/15,875 = 0,567.

На основании полученных соотношений выполним их группировку по кварталам (табл. 8.12).

Данные табл. 8.12 свидетельствуют о наличии различий в величине этих соотношений по кварталам разных лет, хотя их порядок примерно соблюдается. Для расчета индекса сезонности на основании сравнений фактических квартальных значений за ряд лет с соответствующей скользящей средней можно воспользоваться следующими приемами:

1) рассчитать для каждого квартала среднюю арифметическую из полученных соотношений; например, для I квартала среднее значение индекса сезонности составит:

$$\left(\frac{0,960 + 0,965 + 0,895 + 0,929}{4}\right) = 0,9373;$$

2) определить медиану из значений индексов сезонности за каждый квартал. Например, проранжировав значения соотношений I квартала, получаем следующий ряд индексов: 0,895; 0,929; 0,960; 0,965. Тогда медиана будет равна средней величине из второго и третьего значений: Me = (0,929 + 0,960)/2 = 0,9445;

3) определить так называемую модифицированную среднюю. Величина этой средней находится путем исключения из рассмотре-

Таблица 8.12

Соотношения между фактическим объемом товарооборота и скользящей средней

Год	Кварталы			
	I	II	III	IV
1991	—	—	0,567	1,152
1992	0,960	1,270	0,625	1.075
1993	0,965	1,297	0,684	1,150
1994	0,895	1,342	0,591	1,113
1995	0,929	1,284	—	—

ния по каждому кварталу наименьшего и наибольшего значений и определения средней арифметической из оставшихся значений индексов сезонности. Для расчета модифицированной средней первого квартала исключаем значения 0,895 и 0,965. Тогда

$$\bar{x}_{\text{модиф}} = \frac{0{,}929 + 0{,}960}{2} = 0{,}9445.$$

Поскольку в рассматриваемом примере для каждого квартала мы располагаем четырьмя значениями, расчет медианы приводит к такому же результату, что и расчет модифицированной средней. Поэтому в табл. 8.13 приведены только средние арифметические значения индексов и медианы.

В итоговой строке таблицы (см. графы 2 и 3) сумма индексов сезонности хотя и незначительно, но отличается от 4 (для четырех кварталов сумма индексов должна быть равна 4, а их средняя равна 1). Для устранения этих расхождений определяется поправочный коэффициент как отношение теоретической суммы индексов (4) к фактической величине их суммы (3,9666 или 3,9745). Результаты расчета поправочного коэффициента приведены в последней строке табл. 8.13. Воспользуемся значениями медианы для расчета индексов сезонности, скорректированных на поправочный коэффициент. Так, для первого квартала скорректированное значение медианы составит 0,95056 (0,9445×1,0064). Аналогичным образом рассчитаны значения и для остальных кварталов (см. графу 4 табл. 8.13). Округлив до четвертого знака после запятой, получим значение индексов сезонности (см. графу 7 табл. 8.11).

Прежде чем анализировать основную тенденцию (тренд) или циклические колебания, необходимо исключить сезонный компо-

Таблица 8.13

Расчет индексов сезонности

Кварталы	Средний арифметический индекс сезонности	Медиана	Скорректированное значение медианы
1	2	3	4
I	0,9373	0,9445	0,95056
II	1,2987	1,2905	1,29876
III	0,6168	0,6080	0,61189
IV	1,1138	1,1315	1,13874
Итого	3,9666	3,9745	3,99995 — 4,00
Поправочный коэффициент	1,0084	1,0064	

нент. Без учета сезонности объем квартального товарооборота представлен в графе 8 табл. 8.11. Эти величины получены путем деления фактического объема товарооборота на индекс сезонности. Так, для первого квартала 1994 г. получим 16 : 0,9506 = 16,8 тыс. долл., для второго квартала 1994 г. — 21 : 1,2988= 16,2 и т.д. После исключения сезонной составляющей из колеблемости уровней временного ряда рассчитаем уравнение тренда, воспользовавшись линейной функцией $\bar{y}_t = a + bt$.

В табл. 8.14 представлен расчет параметров линейного уравнения тренда. Напомним, что при условии $\sum t_i = 0$,

$$a = \frac{\sum_{i=1}^{n} y_i}{n}, \; b = \frac{\sum_{i=1}^{n} y_i t_i}{\sum_{i=1}^{n} t_i^2},$$

где $n = 20$

Используя итоги граф 3, 5 и 6 табл. 8.14, находим:

$$a = \frac{360{,}9}{20} = 18{,}045; \; b = \frac{420{,}3}{2660} = 0{,}158.$$

Следовательно, уравнение тренда для данного примера имеет следующий вид:

$$\bar{y}_t = 18{,}045 + 0{,}158t.$$

С помощью полученного уравнения тренда выполним экстраполяцию на I и II кварталы 1999 г. Подставив в уравнение $t = 21$, получим $\bar{y}_t = 21{,}36$ тыс. долл., при $t = 23$ $\bar{y}_t = 21{,}68$ тыс. долл.

Найденные таким образом значения не учитывают сезонные колебания в объеме товарооборота. Для учета сезонной составляющей уровень, полученный в результате экстраполяции, умножают на индекс сезонности:

$$\bar{y}^1 = \bar{y}_t \times i_{\text{сез}},$$

где $\bar{y}^1$ — экстраполируемый уровень с учетом сезонных колебаний.

Тогда для I квартала 1999 г. получим 20,30 тыс. долл. (21,36×0,9506), а для II квартала 1999 г. — 28,16 тыс. долл. (21,68×1,2988).

Выравненные по уравнению тренда значения товарооборота, скорректированные с учетом сезонности, представлены в графе 8 табл. 8.14. При сопоставлении рассчитанных значений с фактическим объемом товарооборота по кварталам, приведенным в графе 3

Таблица 8.14

Год	Квартал	Объем товарооборота без учета сезонности (в тыс. долл.) y_i	Условные обозначения периодов t_i	t_i^2	$y_i t_i$	$\tilde{y}_i$	$\hat{y}_i$, скорректированный с учетом сезонности $\tilde{y}_i \times i_{сез}$
1	2	3	4	5	6	7	8
1994	I	16,8	−19	361	−319,2	15,04	14,30
	II	16,2	−17	289	−275,4	15,36	19,95
	III	14,7	−15	225	−220,5	15,68	9,59
	IV	15,8	−13	169	−205,4	15,99	18,21
1995	I	15,8	−11	121	−173,8	16,31	15,50
	II	15,4	−9	81	−138,6	16,62	21,59
	III	16,3	−7	49	−114,1	16,94	10,37
	IV	15,8	−5	25	−79,0	17,26	19,65
1996	I	17,9	−3	9	−53,7	17,57	16,70
	II	18,5	−1	1	−18,5	17,89	23,24
	III	21,2	+1	1	+21,2	18,20	11,14
	IV	19,3	+3	9	57,9	18,52	21,09
1997	I	17,9	+5	25	89,5	18,84	17,91
	II	19,2	+7	49	134,4	19,15	24,87
	III	18,0	+9	81	162,0	19,47	11,91
	IV	18,4	+11	121	202,4	19,78	22,52
1998	I	18,9	+13	169	245,7	20,10	19,11
	II	20,0	+15	225	300,0	20,42	26,52
	III	22,9	+17	289	389,3	20,73	12,69
	IV	21,9	+19	361	416,1	21,05	23,97
Итого		360,9		2660	420,3		

табл. 8.14, можно видеть, что величина расхождений между ними невелика (максимальное расхождение составляет 10,62%).

Следует подчеркнуть, что величины товарооборота, полученные в результате экстраполяции, не учитывают наличия отклонений фактических уровней от уровней, рассчитанных по уравнению тренда (сравните значения товарооборота в графах 3 и 7 табл. 8.14).



Рис. 8.6. Динамика товарооборота специализированного магазина зарубежной фирмы за 1994—1999 гг.

Как было показано выше, для прогноза должна быть определена средняя квадратическая ошибка уравнения тренда, с помощью которой для заданного уровня значимости определяют доверительные интервалы прогноза. Графическое представление фактических данных объема товарооборота, скользящих средних и уравнения тренда дано на рис. 8.6.

8.8. Какова аналитическая модель проявления сезонных колебаний?

Выделение сезонной волны можно выполнить на основе построения аналитической модели проявления сезонных колебаний. Построение аналитической модели выявляет основной закон колеблемости данного временного ряда в связи с переходом от периода к периоду и дает среднюю характеристику внутригодовых колебаний.

При исследовании явлений периодического типа в качестве аналитической формы развития во времени принимается уравнение следующего типа (ряд Фурье):

$$y_t = a_0 + \sum\left(a_k \cdot \cos(kt) + b_k \cdot \sin(kt)\right). \qquad (8.20)$$

В этом уравнении величина k определяет гармонику ряда Фурье и может быть взята с разной степенью точности (чаще всего от 1 до 4).

Покажем использование ряда Фурье для выявления характера изменения индекса Доу—Джонса (*DJIA*) на основе данных по результатам торгов на Нью-Йоркской бирже в период с 1 июля 1999 г. по 18 ноября 1999 г. (Приложение 10).

Весь период наблюдения был разбит на 20 недель, после чего исходный временной ряд приобрел следующий вид:

Неделя	*DJIA*	Неделя	*DJIA*
1	11 145,40	11	10 856,69
2	11 180,80	12	10 648,33
3	11 061,32	13	10 280,30
4	10 871,05	14	10 515,39
5	10 701,22	15	10 320,74
6	10 782,74	16	10 296,30
7	11 043,94	17	10 479,87
8	11 239,54	18	10 650,40
9	10 856,85	19	10 644,84
10	11 061,32	20	10 903,04

По этим данным построен график динамики индекса за весь рассматриваемый период (рис. 8.7). На графике видно, что величина индекса изменяется волнообразно, поэтому для характеристики этих колебаний в качестве аналитической формы изменения во времени индекса применим ряд Фурье.

Для отыскания параметров уравнения воспользуемся методом наименьших квадратов. После некоторых преобразований получаем систему нормальных уравнений, позволяющих определить необходимые параметры уравнения ряда Фурье:

$$a_0 = \frac{1}{n}\sum_{1}^{n} y_i ;$$

$$a_k = \frac{2}{n}\sum_{1}^{n} y_i \cdot \cos(kt_i) ;$$

$$b_k = \frac{2}{n}\sum_{1}^{n} y_i \cdot \sin(kt_i) .$$

Для изучения сезонных колебаний на протяжении 20 недель необходимо взять $n = 20$. Момент времени t для каждого значения индекса определяется по формуле $\frac{2\pi}{n} \cdot t$. Таким образом, для данного ряда динамики получаем, что первая неделя соответствует $t = 0$, вторая — $t = \frac{\pi}{10}$ и т.д.

В табл. 8.15 показаны все необходимые расчеты для нахождения параметров для построения модели сезонной волны (расчеты для гармоник ряда Фурье с 1 по 4).

После вычисления параметров ряда Фурье были построены графики аналитически выравненных значений индекса *DJIA* (рис. 8.7), из которых видно, что наиболее близкие значения дает ряд Фурье с гармоникой, равной 4. Таким образом, уравнение ряда Фурье для нашего случая имеет вид:

$$\bar{y}_t = 10\,777{,}02 + 91{,}12 \cdot \cos(t) + 294{,}64 \cdot \sin(t) + 179{,}87 \cdot \cos(2t) -$$

$$-100{,}53 \cdot \sin(2t) + 10{,}33 \cdot \cos(3t) + 75{,}39 \cdot \sin(3t) + 8{,}89 \cdot \cos(4t) +$$

$$+55{,}07 \cdot \sin(4t).$$

График наглядно показывает достаточную степень близости фактических и выравненных значений.

8.9. Что такое модели авторегрессии?

Наличие зависимости между последующими и предшествующими уровнями динамического ряда называют ***автокорреляцией***, а построение модели зависимости будущих значений рассматриваемого показателя от прошлых его значений носит название ***авторегрессии***. В моделях авторегрессии предшествующие значения временного ряда рассматриваются как независимые переменные, а по аналогии с уравнениями регрессии их можно назвать факторными признаками.

Если текущие значения показателя рассматривать как функцию от нескольких m предшествующих значений, то общий вид линейной модели авторегрессии можно записать следующим образом:

$$\hat{y}_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \ldots + a_m y_{t-m}.$$

Таблица 8.15

Параметры ряда Фурье для индекса Доу—Джонса

Неделя	t	A_1^1	B_1^1	A_2^1	B_2^1	A_3^1	B_3^1	A_4^1	B_4^1
0	0	1 114,54	0,00	1 114,54	0,00	1 114,54	0,00	1 114,54	0,00
1	0,314159	1 063,36	345,51	904,55	657,19	657,19	904,55	345,51	1063,36
2	0,628319	894,90	650,19	341,82	1052,02	−341,82	1052,02	−894,90	650,19
3	0,942478	638,98	879,49	−335,93	1033,90	−1033,90	335,93	−879,49	−638,98
4	1,256637	330,69	1017,75	−865,75	629,00	−865,75	−629,00	330,69	1017,75
5	1,570796	0,00	1078,27	−1078,27	0,00	−0,00	−1078,27	1078,27	−0,00
6	1,884956	−341,28	1050,34	−893,47	−649,15	893,47	−649,15	341,82	1050,34
7	1,884956	−660,64	909,30	−347,32	−1068,94	1068,94	347,32	−909,30	660,64
8	2,199115	−878,34	638,15	335,50	−1032,55	335,50	1032,55	−878,34	−638,15
9	2,513274	−1 051,99	341,81	894,88	−650,17	−650,17	894,88	341,81	−1051,99
10	2,827433	−1 085,67	0,00	1085,67	−0,00	−1085,67	0,00	1085,67	−0,00
11	3,141593	−1 012,72	−329,05	861,47	625,89	−625,89	−861,47	329,05	1 012,72
12	3,455752	−831,69	−604,26	317,68	977,71	317,68	−977,71	−831,69	604,26
13	3,769911	−618,08	−850,71	−324,94	1000,07	1000,07	−324,94	−850,71	−618,08
14	4,08407	−318,93	−981,56	−834,97	606,64	834,97	606,64	318,93	−981,56
15	4,712389	−0,00	−1029,63	−1029,63	0,00	0,00	1029,63	1029,63	−0,00
16	5,026548	323,85	−996,69	−847,84	−615,99	847,84	615,99	323,85	996,69
17	5,340708	626,01	−861,64	−329,12	−1012,91	−1012,91	−329,12	−861,64	626,01
18	5,654867	861,19	−625,69	328,94	−1012,38	−328,94	−1012,38	−861,19	−625,69
19	5,969026	1 036,94	−336,92	882,07	−640,86	640,86	−882,07	336,92	−1036,94
Итого		91,12	294,64	179,87	−100,53	70,33	75,39	8,89	55,07
		a_1	b_1	a_2	b_2	a_3	b_3	a_4	b_4

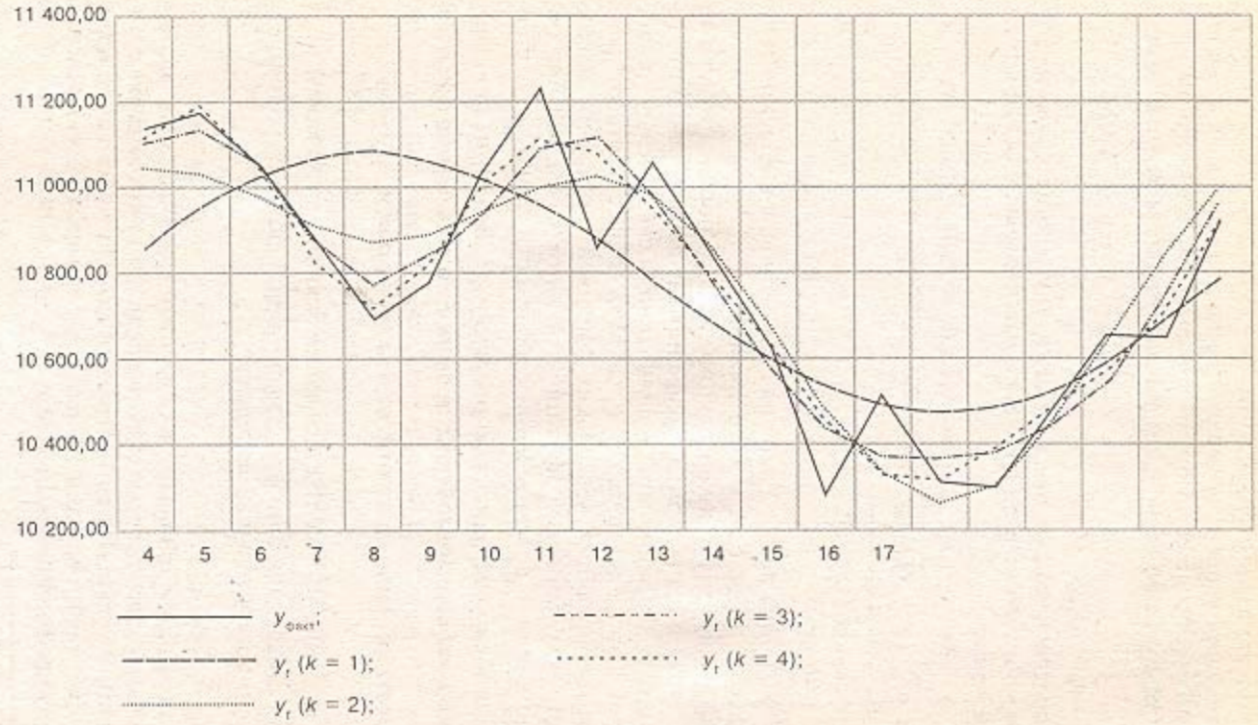


Рис. 8.7. Динамика индекса Доу—Джонса за период 01.07 — 18.11.99 г.

Величина m называется *порядком авторегрессии*.

Если текущий уровень временного ряда зависит лишь от непосредственно ему предшествующего значения, то мы получим уравнение авторегрессии первого порядка, которое имеет следующий вид:

$$\tilde{y}_t = a_0 + a_1 y_{t-1}.$$

Для определения наличия связи между последовательными уровнями одного и того же ряда динамики рассчитывают коэффициент автокорреляции:

$$r_m = \frac{\sum_{t=1}^{n-m} \left(y_t - \bar{y}\right)\left(y_{t+m} - \bar{y}\right)}{\sum_{t=1}^{n} \left(y_t - \bar{y}\right)^2},$$

где m — величина лага, т.е. количество периодов смещения текущего и предшествующего уровня.

Так, коэффициент автокорреляции первого порядка будет рассчитан с лагом в один период, коэффициент автокорреляции второго порядка будет оценивать степень тесноты связи между уровнями, отстоящими на два периода, и т.д. Можно рассчитать коэффициенты автокорреляции нескольких порядков и провести оценку значимости полученных коэффициентов автокорреляции при определенном уровне доверительной вероятности с помощью стандартной ошибки коэффициента корреляции ρ_r. Величина стандартной ошибки, используемой при проверке значимости коэффициентов автокорреляции, рассчитывается по формуле: $S_r = \frac{1}{\sqrt{n}}$.

Зная значение t_α, соответствующее уровню значимости α, и величину стандартной ошибки коэффициента автокорреляции, определим его доверительный интервал[26]:

$$-t_\alpha \cdot S_r \le r_m \le t_\alpha \cdot S_r.$$

Если значение r_m попадает в указанный интервал, то с данной доверительной вероятностью можно предполагать отсутствие автокорреляции. В уравнение авторегрессии следует включать те предшествующие уровни временного ряда, для которых коэффициенты автокорреляции оказались значимыми.

[26] Коэффициенты автокорреляции случайных данных имеют выборочное распределение, близкое к нормальному с нулевым математическим ожиданием и средним квадратическим отклонением, равным $1/\sqrt{n}$.

Глава 9
ИНДЕКСЫ И ИХ ИСПОЛЬЗОВАНИЕ В ЭКОНОМИКО-СТАТИСТИЧЕСКИХ ИССЛЕДОВАНИЯХ

9.1. Что такое индексы и каково значение индексного метода анализа?

В практике статистики индексы наряду со средними величинами являются наиболее распространенными статистическими показателями. С их помощью характеризуется развитие национальной экономики в целом и ее отдельных отраслей, анализируются результаты производственно-хозяйственной деятельности предприятий и организаций, исследуется роль отдельных факторов в формировании важнейших экономических показателей, выявляются резервы производства, индексы используются также в международных сопоставлениях экономических показателей, определении уровня жизни, мониторинге деловой активности в экономике и т.д.

Индекс[27] представляет собой относительную величину, получаемую в результате сопоставления уровней сложных социально-экономических показателей во времени, пространстве или с планом.

В развитии индексной теории в нашей стране сложились два направления: обобщающее, или синтетическое, и аналитическое. Различие между этими направлениями обусловлено двумя возможностями интерпретации индексов.

Обобщающее, или так называемое ***синтетическое, направление*** трактует индекс как показатель среднего изменения уровня изучаемого показателя. В аналитической теории индексы — это показатели изменения уровня результативной величины под влиянием изменения индексируемой величины[28].

Например, величина индекса цен продукции, равная 1,1536, может быть истолкована в обобщающем значении как величина, указывающая на то, что цены на продукцию возросли в целом на 15,36%, или в 1,1536 раза, а в аналитическом значении как показатель того, что в связи с изменением цен стоимость продукции (или

[27] Слово «индекс» (index) буквально означает указатель, показатель.

[28] Величина, изменение которой изучается в данном случае с помощью индекса, называется индексируемой величиной.

размер выручки) увеличилась на 15,36%. Развитие второго направления было обусловлено применением индексного метода в экономическом анализе. Аналитическое направление в развитии индексной теории в значительной степени сформировалось благодаря работам российских статистиков В.Н. Старовского, Н.М. Виноградовой, Л.В. Некраша, И.Ю. Писарева и др. В настоящее время индексный метод продолжает успешно развиваться в трудах многих отечественных статистиков и получил широкое применение в практике статистической работы.

Таким образом, с помощью индексных показателей решаются следующие основные задачи:

1) характеристика общего изменения сложного экономического показателя (например, затрат на производство продукции, стоимости произведенной продукции и т.д.) или формирующих его отдельных показателей-факторов;

2) выделение в изменении сложного показателя влияния одного из факторов путем элиминирования влияния других факторов (например, увеличение выручки от реализации продукции, связанное с ростом цен или выпуска продукции в натуральном выражении). В качестве самостоятельной можно выделить задачу обособления влияния изменения структуры явления на индексируемую величину (например, при изучении динамики среднеотраслевой себестоимости продукции исследуется влияние изменений в распределении объемов выпуска продукции по предприятиям отрасли).

Способы построения индексов зависят от содержания изучаемых показателей, методологии расчета исходных статистических показателей, имеющихся в распоряжении исследователя статистических данных и целей исследования.

По степени охвата элементов совокупности различают индивидуальные и сводные (общие) индексы. ***Индивидуальными*** называются индексы, характеризующие изменение только одного элемента совокупности (например, изменение выпуска легковых автомобилей определенной марки). Индивидуальный индекс обозначается буквой *i*. ***Сводный индекс*** отражает изменение по всей совокупности элементов сложного явления. Если индексы охватывают не все элементы сложного явления, а лишь часть, то их называют ***групповыми*** или субиндексами. Например, сводный индекс характеризует динамику объема промышленной продукции. К субиндексам в этом случае могут быть отнесены индексы продукции по отдельным отраслям промышленности. Обозначают сводный (общий) индекс буквой *I*.

В зависимости от содержания и характера индексируемой величины различают индексы ***количественных*** (объемных) показателей (например, индекс физического объема продукции) и индексы ***качественных*** показателей (например, индексы цен, себестоимости).

При вычислении индексов различают ***сравниваемый уровень*** и уровень, с которым производится сравнение, называемый ***базисным***. Выбор базы сравнения определяется целью исследования. В индексах, характеризующих изменение индексируемой величины во времени, за базисную величину принимают размер показателя в каком-либо периоде, предшествующем отчетному. При этом возможны два способа расчета индексов — цепной и базисный. ***Цепные индексы*** получают путем сопоставления текущих уровней с предшествующим. Таким образом, база сравнения непрерывно меняется. Базисные индексы получают путем сопоставления с уровнем периода, принятого за базу сравнения.

При территориальных сравнениях за базу принимают данные по какой-либо одной части территории (например, при региональных сопоставлениях внутри России) или итоговый показатель по всей изучаемой территории в целом, как это имеет место в международных сопоставлениях.

При использовании индексов как показателей выполнения плана за базу сравнения принимаются плановые показатели.

В зависимости от методологии расчета различают ***агрегатные индексы*** и ***средние из индивидуальных индексов***. Последние, в свою очередь, делятся на средние арифметические и средние гармонические индексы.

Агрегатные индексы качественных показателей могут быть рассчитаны как ***индексы переменного состава*** и ***индексы фиксированного (постоянного) состава***. В индексах переменного состава сопоставляются показатели, рассчитанные на базе изменяющихся структур явлений, а в индексах фиксированного состава — на базе неизменной структуры явлений.

9.2. На каких принципах базируется расчет агрегатных индексов количественных показателей?

Необходимость в применении особых приемов построения индексов количественных показателей возникает, когда итоги по отдельным элементам сложного явления непосредственно несоизмеримы. Например, предприятие экспортирует станки, металл, товары широкого потребления. Если имеются сведения об экспорте продукции только в натуральном выражении, то динамику экспорта продукции предприятия в целом нельзя охарактеризовать показа-

телем $\frac{\sum_{i=1}^{n} q_i^1}{\sum_{i=1}^{n} q_i^0}$, где q_i^1 — количество продукции данного вида в нату-

ральном выражении, экспортируемой в отчетном периоде; q_i^0 — количество продукции того же вида, отправленной на экспорт в базисном периоде.

Различные виды продукции неравноценны по количеству затраченного на них общественного труда и имеют разные потребительские стоимости. Совсем не безразлично, увеличился ли экспорт станков или экспорт детских игрушек на одно и то же количество штук. Поэтому было бы неправильно непосредственно суммировать итоги по разным видам продукции. Для получения общего итога необходимо данные по различным видам продукции привести к единой, общей мере, например, использовать стоимостную оценку экспорта продукции. Тогда вместо $\sum_{i=1}^{n} q_i$ получим сумму вида $\sum_{i=1}^{n} p_i \cdot q_i$, где p_i — цена единицы продукции данного вида (при расчете экспорта это будет внешнеторговая цена). Такой переход от одних единиц измерения к другим в теории индексов называют ***соизмерением***.

При построении индексов объемных показателей в качестве соизмерителей применяют те или иные качественные показатели, например, цену, себестоимость или трудоемкость единицы изделия. Выбор коэффициента соизмерения в каждом конкретном случае зависит от цели исследования. Универсальное значение в индексах физического объема имеют цены. Стоимость всей выработанной на предприятии продукции получаем путем умножения на цену количества выпущенной продукции каждого вида и суммирования произведений по всем видам продукции. Тогда стоимость продукции базисного периода будет определена так:

$$p_1^0 q_1^0 + p_2^0 q_2^0 + p_3^0 q_3^0 + \ldots + p_n^0 q_n^0 = \sum_{i=1}^{n} p_i^0 q_i^0 ,$$

а стоимость продукции отчетного периода составит:

$p_1^1 q_1^1 + p_2^1 q_2^1 + p_3^1 q_3^1 + \ldots + p_n^1 q_n^1 = \sum_{i=1}^{n} p_i^1 q_i^1$, где q_i^0, q_i^1 — количество единиц отдельных видов продукции, соответственно в базисном и отчетном периодах; p_i^0, p_i^1 — цена единицы отдельных видов продукции соответственно в базисном и отчетном периодах; i =1, 2, 3,..., n — количество отдельных видов продукции.

Если разделить стоимость продукции отчетного периода на стоимость продукции базисного периода, получим ***индекс стоимости продукции***. В общем виде его можно записать так:

$$I_{pq} = \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^0} . \qquad (9.1)$$

Если стоимость продукции в отчетном периоде выросла, то индекс стоимости будет больше 1, а если она снизилась, то I_{pq} будет меньше 1.

Приведенная формула характеризует изменение стоимости продукции, которая зависит от изменения уровня цен и количества выпускаемой продукции в отчетном периоде по сравнению с базисным. То есть индекс стоимости показывает ее изменение в результате совместного влияния двух факторов, поэтому индекс стоимости не дает количественного представления об изменении объема выпуска. Это представление мы получим, если элиминируем влияние изменения цен, для чего количество продукции, произведенной в отчетном и базисном периодах, умножим на одинаковые для обоих периодов цены:

$$I_q = \frac{\sum_{i=1}^{n} q_i^1 p_i}{\sum_{i=1}^{n} q_i^0 p_i} . \qquad (9.2)$$

Такой индекс называют ***агрегатным индексом физического объема***.

При вычислении индекса физического объема продукции возможны разные решения в зависимости от выбора коэффициента соизмерения.

Если принять за коэффициент соизмерения цены базисного периода, то индекс физического объема продукции будет иметь следующий вид:

$$I_q = \frac{\sum_{i=1}^{n} p_i^0 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^0} . \qquad (9.3)$$

(Такой вариант построения агрегатного индекса был предложен Э. Ласпейресом в 1864 г.)

Внешней отличительной особенностью агрегатного индекса является то, что в числителе и знаменателе меняется индексируемая величина, значения же другой, являющейся соизмерителем, остаются неизменными. В приведенном варианте индекса количественных показателей значения соизмерителей принимаются на уровне базисного периода. Используя коэффициенты соизмерения базисного периода, берут базисные соотношения по уровню цен, но зато полностью элиминируют влияние на изменение стоимости продукции изменения самих цен.

В практике планирования при проведении экономико-статистического анализа не ограничиваются исчислением отдельных, изолированных индексов, характеризующих изменение показателя за какой-то один период времени. Исчисляют, как правило, не один индекс, а несколько индексов за последовательные периоды времени. При таком исчислении обычно применяют во всех индексах в качестве соизмерителей цены одного и того же периода, например, для динамических сопоставлений роста выпуска объема продукции в промышленности, строительстве и т.д. Такие цены называются ***сопоставимыми*** (***фиксированными***, или ***неизменными***); в условиях стабильной экономики они применяются на протяжении длительного периода времени. При существенных различиях в соотношении уровней действующих и фиксированных цен производится пересмотр последних, и они меняются время от времени с изменением особенностей самого ценообразования. В настоящее время в странах СНГ, учитывая нестабильное состояние экономики, при расчетах динамики валового внутреннего продукта, национального богатства в качестве фиксированных используют цены предыдущего года.

Применяя в качестве соизмерителя неизменные цены, получим следующую формулу индекса физического объема продукции:

$$I_q = \frac{\sum_{i=1}^{n} p_i'' q_i^1}{\sum_{i=1}^{n} p_i'' q_i^0}, \qquad (9.3а)$$

где p_i'' — неизменная оптовая цена единицы изделия.

Преимущество такого варианта соизмерения продукции состоит и в том, что путем суммирования может быть получен итоговый показатель за период любой продолжительности, т.е. на основе данных о стоимости продукции за каждый месяц можно получить стоимость продукции за квартал, полугодие, год. Использование неизменных цен в учете продукции дает возможность изучать динамику выпуска не только отдельных видов продукции, но и по предприятиям, отраслям промышленности и промышленности в целом.

Если поставить задачу характеристики изменения физического объема выпуска продукции, то правомерно в качестве соизмерителей использовать и цены отчетного периода, тогда индекс физического объема продукции будет записан так:

$$I_q = \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^1 q_i^0}. \qquad (9.4)$$

Агрегатный индекс с соизмерителями отчетного периода был предложен в 1874 г. Г. Пааше.

Рассмотрим порядок вычисления индексов физического объема продукции на следующем условном примере. Имеются данные (табл. 9.1) о фактическом выпуске продукции машиностроительным предприятием за два года. По данным о выпуске продукции в натуральном выражении можно рассчитать индексы, характеризующие динамику выпуска отдельных видов продукции, или индивидуальные индексы. ***Индивидуальный индекс*** показывает, во сколько раз изменилось производство данного вида продукции в отчетном периоде по отношению к периоду, с которым проводилось сравнение.

Для вычисления индивидуальных индексов динамики определяют отношение объема выпуска продукции отчетного периода к объему выпуска в предшествующем периоде (см. графу 6 табл. 9.1):

Таблица 9.1

Объем выпуска продукции предприятия по видам

Виды продукции (в тыс. руб.)	Выпуск продукции в натуральном выражении		Цена производителя за единицу (в млн руб.)		Индивидуальные индексы физического объема продукции $i_q = q_i^1/q_i^0$	Индивидуальные индексы цен $i_p = p_i^1/p_i^0$
	базисный период q_i^0	отчетный период q_i^1	базисный период p_i^0	отчетный период p_i^1		
1	2	3	4	5	6	7
Оборудование	2100	2000	75,00	82,50	0,9524	1,1000
Литье (в тоннах)	11 500	12 000	8,75	10,10	1,0435	1,1543

$$i_q = \frac{q_i^1}{q_i^0}.$$

Индекс динамики объема производства оборудования составляет 95,24%, что означает снижение его выпуска на 4,76% ($i_q \times 100\% - 100\%$). В динамике же выпуска литья наблюдается противоположная тенденция: выпуск литья возрос на 4,35% ($1{,}0435 \times 100\% - 100\%$). Общее изменение выпуска продукции предприятия может быть получено на основе определения агрегатной формы индекса физического объема продукции. Покажем расчет агрегатных индексов физического объема продукции в двух вариантах:

1) в качестве соизмерителей используются цены базисного периода (формула (9.3));

2) соизмерителями разнородной продукций предприятия являются текущие цены (цены отчетного периода) (формула (9.4)).

Стоимостные показатели выпуска продукции, необходимые для расчета индексов, приведены в табл. 9.2.

Агрегатный индекс[29] динамики физического объема продукции, рассчитанный по формуле Ласпейреса, составит 0,9879:

$$I_q^{Л} = \frac{\sum_{i=1}^{n} p_i^0 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^0} = \frac{255\,000}{258\,125} = 0{,}9879, \text{ или } 98{,}79\%,$$

Таблица 9.2

Расчет стоимости выпуска продукции

Виды продукции	Стоимость выпуска (в тыс. руб.)		Условная стоимость выпуска (в тыс. руб.)		Доля изделий в стоимости продукции предприятий	
	базисный период $p_i^0 q_i^0$	отчетный период $p_i^1 q_i^1$	базисный период в текущих ценах $p_i^1 q_i^0$	отчетный период в базисных ценах $p_i^0 q_i^1$	базисный период d_i^0	отчетный период d_i^1
1	2	3	4	5	6	7
Оборудование	157 500	165 000	173 250	150 000	0,6102	0,5765
Литье	100 625	121 200	116 150	105 000	0,3898	0,4235
Итого	258 125	286 200	289 400	255 000	1,0000	1,0000

[29] Индексы могут быть выражены в долях единицы или процентах.

где $I_q^{Л}$ - агрегатный индекс физического объема продукции, рассчитанный по формуле Ласпейреса.

Таким образом, физический объем выпуска продукции предприятия снизился в отчетном периоде на 1,21%.

Величина агрегатного индекса физического объема, рассчитанная по формуле Пааше $I_q^{П}$, равна 0,98894:

$$I_q^{П} = \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^1 q_i^0} = \frac{286\,200}{289\,400} = 0{,}98894,$$

т.е. физический объем выпуска продукции предприятия уменьшился на 1,106%.

Если сопоставить величины двух индексов $I_q^{Л}$ и $I_q^{П}$, то, несмотря на некоторые различия в величине, они отражают одну и ту же тенденцию — снижение физического объема выпуска продукции предприятия.

Какой бы вариант построения агрегатного индекса ни был выбран — индекс Ласпейреса или индекс Пааше, его необходимо использовать постоянно, иначе сравнение значений индекса в разные периоды времени будет бессмысленным.

9.3. Какие виды средних индексов можно использовать для обобщения динамики количественных показателей?

Величина агрегатного индекса физического объема зависит от индивидуальных индексов, так как общее изменение объема производимой продукции (при неизменности ассортимента) есть результат изменения объема выпуска каждого отдельного вида. Общий результат изменения определяется также по удельному весу стоимости отдельных видов продукции в общей стоимости продукции.

Общий индекс физического объема, построенный на базе индивидуальных индексов, принимает форму среднего арифметического или гармонического индекса. Например, известна стоимость продукции каждого вида в базисном периоде ($p_i^0 q_i^0$) и индивидуальные индексы физического объема $i_q = \frac{q_i^1}{q_i^0}$. Исходной базой построения среднего из индивидуальных индексов служит агрегатная форма индекса Ласпейреса:

$$I_q = \frac{\sum_{i=1}^{n} p_i^0 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^0}.$$

Из имеющихся данных непосредственно путем суммирования можно получить только знаменатель формулы. Числитель же может быть получен путем перемножения стоимости отдельного вида продукции базисного периода на индивидуальный индекс:

$$q_i^1 p_i^0 = q_i^0 p_i^0 \cdot i_{q_i} = q_i^0 p_i^0 \cdot \frac{q_i^1}{q_i^0}.$$

Тогда формула агрегатного индекса физического объема принимает вид:

$$I_q = \frac{\sum_{i=1}^{n} i_{q_i} p_i^0 q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0}, \qquad (9.5)$$

т.е. мы получили средний арифметический индекс физического объема, где весами служит стоимость отдельных видов продукции в базисном периоде.

При выборе весов следует иметь в виду, что средний индекс должен быть тождествен агрегатному, который является основной формой индекса. Учитывая, что отношение $\frac{p_i^0 q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0}$ характеризует долю данного вида продукции в общей стоимости продукции базисного периода (d_i^0), средний арифметический индекс физического объема будет иметь вид:

$$I_q = \sum_{i=0}^{n} i_{q_i} d_i^0. \qquad (9.5а)$$

Воспользуемся данными графы 6 табл. 9.1 и графы 2 табл. 9.2 для расчета среднего арифметического индекса физического объема продукции:

$$I_q = \frac{0{,}9524 \cdot 157\,500 + 1{,}0435 \cdot 100\,625}{157\,500 + 100\,625} = 0{,}9879$$

и получим такой же результат, как и при расчете агрегатного индекса физического объема по формуле Ласпейреса.

Снижение общего объема выпуска продукции предприятия на 1,21% объясняется тем, что превалирующее влияние на величину агрегатного индекса оказывает изменение физического объема выпуска оборудования, поскольку доля оборудования в стоимости продукции предприятия в базисном периоде составляла 61,02% (см. графу 6 табл. 9.2).

Допустим, что имеется информация о динамике объема выпуска каждого вида продукции (I_q) и стоимости каждого вида продукции в отчетном периоде ($p_i^1 q_i^1$).

Для определения общего изменения выпуска продукции предприятия в этом случае удобно воспользоваться формулой Пааше, так как числитель формулы можно получить путем суммирования величин $p_i^1 q_i^1$, а знаменатель — путем деления фактической стоимости каждого вида продукции на соответствующий индивидуальный индекс физического объема продукции, т.е. путем деления:

$$\frac{p_i^1 q_i^1}{i_q}\left(\frac{p_i^1 q_i^1}{\frac{q_i^1}{q_i^0}} = p_i^1 q_i^0\right),$$

$$\text{тогда } I_q^{\Pi} = \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^1 q_i^0} = \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} \frac{p_i^1 q_i^1}{i_{q_i}}}.$$

Таким образом, в этом случае расчет выполняется по формуле среднего взвешенного гармонического индекса физического объема и величина его будет равна 0,98896 (используем данные графы 6 табл. 9.1 и графы 3 табл. 9.2):

$$I_q = \frac{165\,000 + 121\,200}{\frac{165\,000}{0{,}9524} + \frac{121\,200}{1{,}0435}} = 0{,}98896.$$

Можно использовать также данные о стоимости продукции в неизменных (или базисных) ценах (см. графу 5 табл. 9.2). В этом случае, располагая индивидуальными индексами физического объема продукции и величинами $p_i^0 q_i^1$, расчет общего изменения физического объема продукции производим по формуле среднего взвешенного гармонического индекса физического объема продукции:

$$I_q = \frac{\sum_{i=1}^{n} p_i^0 q_i^1}{\sum_{i=1}^{n} \frac{p_i^0 q_i^1}{i_{q_i}}}. \qquad (9.6а)$$

Из сказанного следует, что применение той или иной формулы индекса физического объема (агрегатного, среднего арифметического и среднего гармонического) зависит от имеющейся в нашем распоряжении информации.

И наконец, расчет агрегатных индексов физического объема может производиться на основе данных о стоимостных (а не натуральных) объемах выпуска каждого вида продукции и индивидуальных индексах цен. В условиях рыночной экономики мониторинг цен имеет первостепенное значение.

Допустим, что мы располагаем данными о стоимости выпуска продукции в отчетном периоде $p_i^1 q_i^1$ и базисном периоде $p_i^0 q_i^0$ и индивидуальными индексами цен по отдельным видам продукции:

$$i_{p_i} = \frac{p_i^1}{p_i^0}.$$

Если воспользоваться формулой агрегатного индекса физического объема Ласпейреса, то следует рассчитать числитель $p^0 q^1$ путем деления стоимости продукции отчетного периода на индекс цен.

Тогда получим следующую формулу:

$$I_q^{\text{Л}} = \frac{\sum_{i=1}^{n}\left(p_i^1 q_i^1 : \frac{p_i^1}{p_i^0}\right)}{\sum_{i=1}^{n} p_i^0 q_i^0} = \frac{\sum_{i=1}^{n} \frac{p_i^1 q_i^1}{i_{p_i}}}{\sum_{i=1}^{n} p_i^0 q_i^0}. \qquad (9.7)$$

Если в расчетах динамики выпуска продукции предприятия опираться на индекс Пааше, то следует произвести перерасчет знаменателя формулы путем умножения стоимости продукции базисного периода на индекс цен, т.е. рассчитать величины $p_i^0 q_i^0 \cdot \frac{p_i^1}{p_i^0} = p_i^1 q_i^0$.

В этом случае формула общего индекса физического объема продукции имеет следующий вид:

$$I_q^{\text{П}} = \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n}\left(p_i^0 q_i^0 \cdot \frac{p_i^1}{p_i^0}\right)} = \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^0 \cdot i_{p_i}}. \qquad (9.8)$$

Воспользуемся данными графы 7 табл. 9.1 и граф 2 и 3 табл. 9.2 для определения индексов физического объема продукции. На основе известных индивидуальных индексов цен и стоимостных объемов выпуска продукции получим следующие величины:

$$I_1 = \frac{\frac{165\,000}{1{,}10} + \frac{121\,200}{1{,}1543}}{157\,500 + 100\,625} = \frac{254\,9987{,}7}{258\,125} = 0{,}9879;$$

$$I_2 = \frac{165\,000 + 121\,200}{157\,500 \cdot 1{,}1000 + 100\,625 \cdot 1{,}1543} = \frac{286\,200}{289\,401{,}4} = 0{,}98894.$$

Результаты расчетов аналогичны полученным ранее по формулам агрегатных индексов физического объема Ласпейреса и Пааше.

9.4. В чем состоит различие агрегатных индексов цен и других качественных показателей, рассчитанных по формулам Ласпейреса и Пааше?

Наряду с индексами физического объема продукции в планировании и статистико-экономическом анализе деятельности предприятий и отраслей широко применяются индексы качественных показателей: цен, себестоимости, производительности труда, средней заработной платы и т.д. Качественный показатель характеризует уровень изучаемого результативного показателя в расчете на количественную единицу и определяется как отношение данного результативного показателя к связанному с ним количественному показателю (фактору). Например, себестоимость единицы продукции определяется как отношение суммы затрат на производство этого вида продукции к количеству единиц продукции данного вида; средняя заработная плата определяется путем деления фонда заработной платы на численность работников и т.д.

Рассмотрим формулы индексов качественных показателей на примере расчета индексов цен по данным табл. 9.1.

Индивидуальные индексы цен $i_{p_i} = \frac{p_i^1}{p_i^0}$ характеризуют относительное изменение уровня цен единицы каждого вида продукции в отчетном периоде по сравнению с базисным. Приведенные в графе 7 табл. 9.1 значения индивидуальных индексов цен показывают, что цены на оборудование выросли в 1,10 раза, или на 10%, а цены на литье — в 1,1543 раза, или на 15,43%.

Для определения общего изменения уровня цен на продукцию предприятия, включающую различные виды, нужно рассчитать агрегатный индекс цен. Непосредственное суммирование уровня цен одного станка и одной тонны литья не имеет экономического содержания. Несоизмеримость уровней в таком случае преодолевается путем взвешивания цены каждого вида продукции на количество произведенных единиц, т.е. для отчетного и базисного периодов определяются величины вида $\sum_{i=1}^{n} p_i q_i$, которые и сравниваются между собой. Это сравнение будет отражать только изменение цен, если величина q_i фиксировалась в числителе и знаменателе индекса цен на уровне одного из периодов. Общая формула агрегатного индекса цен записывается так:

$$I_q^{\text{Л}} = \frac{\sum_{i=1}^{n} p_i^1 q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0}. \qquad (9.9)$$

Очевидно, что как и в случае построения агрегатных индексов физического объема, возможен выбор в качестве веса количества продукции отчетного периода q_i^1 (формула Пааше) или количества продукции базисного периода q_i^0 (формула Ласпейреса).

Формула агрегатного индекса цен Ласпейреса:

$$I_q^{\text{Л}'} = \frac{\sum_{i=1}^{n} p_i^1 q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0}. \qquad (9.10)$$

Воспользуемся данными табл. 9.1 и итогами граф 4 и 2 табл. 9.2 для расчета этого индекса:

$$I_q^{\text{Л}} = \frac{82{,}5 \cdot 2100 + 10{,}10 \cdot 11\,500}{75{,}0 \cdot 2100 + 8{,}75 \cdot 11\,500} = \frac{289\,400}{258\,125} = 1{,}1212.$$

Полученная величина индекса означает, что цены на продукцию предприятия возросли в отчетном периоде на 12,12%.

Формула агрегатного индекса цен Пааше:

$$I_q^{\text{П}} = \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^1}. \qquad (9.11)$$

Используя данные табл. 9.1 и 9.2, получим величину агрегатного индекса цен Пааше 1,1224:

$$I_q^{\text{П}} = \frac{82{,}5 \cdot 2000 + 10{,}10 \cdot 12\,000}{75{,}0 \cdot 2000 + 8{,}75 \cdot 12\,000} = \frac{286\,200}{255\,000} = 1{,}1224.$$

По результатам расчета можно констатировать, что цены на всю продукцию предприятия возросли на 12,24%. Значения агрегатных индексов цен, рассчитанные по формулам Ласпейреса и Пааше, будут равны лишь в случае полного совпадения состава продукции отчетного и базисного периодов. Различия в соотношении этих индексов определяются относительной вариацией индивидуальных индексов цен (v_{i_p}), индивидуальных индексов физического объема (v_{i_q}) и коэффициентом корреляции, оценивающим степень тесноты корреляционной связи между названными индивидуальными индексами ($r_{i_p i_q}$):

$$I_q^{\text{Л}} / I_q^{\text{П}} = 1 + r_{i_p i_q} \cdot v_{i_p} \cdot v_{i_q}, \qquad (9.12)$$

где $I_q^{\text{П}}$ и $I_q^{\text{Л}}$ — индексы цен Пааше и Ласпейреса;

v_{i_p} — коэффициент вариации индивидуальных индексов цен:

$$v_{i_p} = \frac{\delta_{i_p}}{I_q^{\text{Л}}}; \quad \delta_{i_p} = \sqrt{\frac{\sum_{i=1}^{n} \left(i_{p_i} - I_q^{\text{Л}}\right)^2 p_i^0 q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0}};$$

v_{i_q} — коэффициент вариации индивидуальных индексов физического объема продукции:

$$v_{i_q} = \frac{\delta_{iq}}{I_q^{\Pi}}; \qquad \delta_{i_q} = \sqrt{\frac{\sum_{i=1}^{n}\left(i_{q_i} - I_q^{\Pi}\right)^2 p_i^0 q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0}}.$$

Степень тесноты связи между индивидуальными индексами цен и физического объема продукции определяется так:

$$r_{i_p i_q} = \frac{\sum\left(i_{p_i} - I_p^{\Pi}\right)\left(i_{q_i} - I_q^{\Pi}\right) p_i^0 q_i^0}{\sum p_i^0 q_i^0 \cdot \delta_{i_p} \cdot \delta_{i_q}}.$$

9.5. Что такое средние взвешенные индексы качественных показателей?

На практике в целом ряде случаев могут быть известны не абсолютные значения индексируемых показателей, а их относительные изменения. Например, могут быть известны изменения уровней цен или себестоимости по отдельным видам продукции, средней заработной платы по отдельным категориям персонала, рентабельности на отдельных предприятиях отрасли и т.д. В таких случаях агрегатный индекс можно рассчитать косвенным путем, используя взвешенную среднюю из индивидуальных индексов, если нам известен размер результативного показателя за отчетный или базисный периоды.

Широко применяется средний взвешенный гармонический индекс в статистике торговли при определении индексов розничных цен. Учет товарооборота ведется в денежном выражении по группам товаров, данные же о количестве проданных товаров в натуральном выражении во многих случаях отсутствуют. Поэтому непосредственно определить условную сумму товарооборота $\sum_{i=1}^{n} p_i^0 q_i^1$ невозможно, и тогда вместо агрегатной формы индекса вычисляется средний гармонический индекс с текущими весами:

$$I_p = \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} \frac{p_i q_i^1}{i_p}}. \qquad (9.13)$$

Он алгебраически тождествен формуле Пааше и имеет точно такое же экономическое содержание.

Рассмотрим вычисление средних взвешенных индексов качественных показателей на примере. За отчетный месяц цена единицы изделия А возросла на 5% по сравнению с предыдущим месяцем, изделия Б — на 3%, изделия В — на 11%. Нужно определить общий (средний) процент роста цен по всем изделиям в отчетном месяце, если известно, что объем товарооборота в отчетном месяце составил: по изделию А — 780 млн руб., по изделию Б — 520 млн руб., по изделию В — 340 млн руб. Имеющиеся данные представлены в табл. 9.3 (графы 3 и 4).

Определяем агрегатный индекс цен (по формуле Пааше):

$$I_q^{\Pi} = \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^1} = \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} \frac{p_i^1 q_i^1}{i_{p_i}}}.$$

Числитель формулы приведен в итоговой строке графы 3 табл. 9.3 и равен 1640 млн руб. Слагаемые знаменателя можно определить

Таблица 9.3

Динамика и структура товарооборота магазина

Изделия	Объем товарооборота (в млн руб.)		Индивидуальные индексы цен i_{p_i}	Условный объем товарооборота (в млн руб.)		Удельный вес стоимости изделия в общем объеме товарооборота	
	в предшествующем месяце $p_i^0 q_i^0$	в отчетном месяце $p_i^1 q_i^1$		в отчетном месяце по ценам предыдущего $(p_i^1 q_i^1)/i_{p_i}$	в предыдущем месяце по ценам отчетного $p_i^0 q_i^0 \times i_{p_i}$	в предыдущем месяце d_i^0	в отчетном месяце d_i^1
1	2	3	4	5	6	7	8
А	750	780	1,05	742,86	787,5	47,02	47,56
Б	530	520	1,03	504,85	545,9	33,23	31,71
В	315	340	1,11	306,31	349,65	19,75	20,73
Итого	1595	1640		1554,02	1683,05	100,00	100,00

путем деления товарооборота данного вида продукции в отчетном году на индивидуальный индекс цен:

$$\frac{p_i^1 q_i^1}{i_{p_i}} = p_i^1 q_i^1 : \frac{p_i^1}{p_i^0} = p_i^0 q_i^1.$$

Так, по изделию А $p_i^0 q_i^1 = 742{,}86$ млн руб. и т. д. (см. графу 5 табл. 9.3).

Таким образом, получен общий объем товарооборота по ценам базисного периода. Общая его сумма $\sum_{i=1}^{n} \frac{p_i^1 q_i^1}{i_{p_i}}$ стоит в знаменателе формулы. Разделив итог графы 3 на итог графы 5, получим, что в среднем цены возросли на 5,53%:

$$I_q^{\text{п}} = \frac{1640}{1554{,}02} = 1{,}0553.$$

В данном случае агрегатный индекс цен представлен в форме среднего гармонического взвешенного индекса. В качестве весов используются фактические объемы товарооборота в отчетном месяце.

Поставим ту же задачу определения общего изменения цен на все изделия, но при условии, что известен товарооборот в предыдущем месяце, т.е. величина $p_i^0 q_i^0$. Тогда при имеющейся информации об индивидуальных индексах цен $i_{p_i} = \frac{p_i^1}{p_i^0}$ и товарообороте предыдущего месяца (данные граф 4 и 2 табл. 9.3) рассчитать общий индекс цен можно с использованием агрегатного индекса Ласпейреса:

$$I_q^{\text{л}} = \frac{\sum_{i=1}^{n} p_i^1 q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0} = \frac{\sum_{i=1}^{n} \left(\frac{p_i^1}{p_i^0} \times p_i^0 q_i^0 \right)}{\sum_{i=1}^{n} p_i^0 q_i^0}. \qquad (9.14)$$

В данном случае агрегатный индекс цен представлен в виде среднего арифметического индекса, а в качестве весов используются фактические объемы товарооборота предыдущего месяца.

Учитывая, что $\frac{p_i^0 q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0}$ представляет собой удельный вес стоимости i-го изделия в общем объеме товарооборота в предыдущем

месяце, в качестве веса могут использоваться и показатели структуры товарооборота предыдущего месяца (см. графу 7 табл. 9.3). Используем итоги граф 6 и 2 табл. 9.3 для расчета агрегатного индекса цен и установим, что в среднем цены возросли на 5,52%:

$$I_q^{\text{л}} = \frac{1683{,}05}{1595} = 1{,}0552.$$

Значения индексов $I_q^{\text{л}}$ и $I_q^{\text{п}}$ достаточно близки по величине, поскольку в структуре товарооборота в отчетном месяце не произошло значительных изменений, хотя нет и полного совпадения структуры товарооборота в текущем и предшествующем месяцах (сравните данные по строкам граф 7 и 8 табл. 9.3). Близость величины агрегатного индекса цен к величине индивидуального индекса цен изделия А объясняется тем, что на долю этого изделия приходится более 47% общего объема товарооборота. «Вклад» отдельных изделий в общий рост цен зависит от структуры товарооборота.

Приведенные варианты исчисления индексов отражают практику отечественной статистики. Во многих странах индексы физического объема и цен также исчисляются аналогичным образом. Вместе с тем в международной статистике для расчетов индексов рекомендуются и другие формы индексов[30].

9.6. Каковы принципиальные основы выбора формы индекса?

Если подходить к принципам построения индексов с формально-математических позиций, то, ориентируясь на принцип элиминирования влияния других факторов, кроме изучаемого, при исчислении индексов можно опираться на веса базисного периода (формула Ласпейреса) или же на веса отчетного периода (формула Пааше). Основываясь на этих двух вариантах построения индексов, Фишер предложил рассчитывать среднюю геометрическую из двух агрегатных индексов, назвав ее «идеальной формулой». В табл. 9.4 представлены варианты расчета агрегатных индексов физического объема и цен, наиболее часто используемых в отечественной и зарубежной практике для характеристики временных или пространственных изменений в уровнях анализируемых показателей.

[30] См., например: Иванов Я. Обзор аксиоматической теории индексов//Вопросы статистики. 1995. № 10. С. 25—39.

Однако обзор далеко не полного перечня теоретически возможных методов построения агрегатных индексов количественных и качественных показателей свидетельствует о многовариантности возможных расчетов этих показателей. Как подходить к выбору наиболее оптимального варианта, что влияет на выбор того или иного варианта индекса? Основные факторы, влияющие на принятие решения по методу расчета агрегатного индекса, определены целями исследования, возможностями получения статистической информации, а также потребностями удешевления затрат на ее получение.

С этих позиций, если ориентироваться на синтетическое направление в использовании индексов, т.е. поставить задачу характеристики общего изменения уровня анализируемого показателя, предпочтение может быть отдано индексу Ласпейреса. Например, при исчислении агрегатного индекса физического объема продукции в этом случае достаточно вести мониторинг за изменением физических объемов продукции, тогда как при использовании варианта агрегатного индекса Пааше должно учитываться изменение и физического объема продукции и цен. Расчет агрегатного индекса физического объема продукции по формуле Ласпейреса получил наибольшее распространение в мировой практике. Опора на неизменную структуру потребления при расчете агрегатного индекса

Таблица 9.4

Название индекса	Агрегатные индексы	
	физического объема	цен
Индекс с базисными весами (формула Ласпейреса)	$\frac{\sum_{i=1}^{n} q_i^1 p_i^0}{\sum_{i=1}^{n} q_i^0 p_i^0}$	$\frac{\sum_{i=1}^{n} p_i^1 q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0}$
Индекс с весами отчетного периода (формула Пааше)	$\frac{\sum_{i=1}^{n} q_i^1 p_i^1}{\sum_{i=1}^{n} q_i^0 p_i^1}$	$\frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^1}$
«Идеальная» формула Фишера	$\frac{\sum_i q_i^1 p_i^0}{\sum_i q_i^0 p_i^0} \times \frac{\sum_i q_i^1 p_i^1}{\sum_i q_i^0 p_i^1}$	$\frac{\sum_i p_i^1 q_i^0}{\sum_i p_i^0 q_i^0} \times \frac{\sum_i p_i^1 q_i^1}{\sum_i p_i^0 q_i^1}$

цен также обусловила применение формулы Ласпейреса при расчете индекса потребительских цен (ИПЦ), величина которого используется при индексации доходов населения (подробнее см. параграф 9.12).

Еще одно преимущество формулы агрегатного индекса Ласпейреса связано с возможностями перехода от ряда индексов с переменной базой сравнения к ряду индексов с постоянной базой сравнения и обратно.

9.7. Какова взаимосвязь цепных и базисных индексов?

Как уже указывалось, для изучения динамики показателя за ряд периодов возможно вычисление системы цепных и базисных индексов. Расчет такой системы индексов осуществляется в двух вариантах:

1) сравнивают размер показателя в различные периоды с уровнем того же показателя в какой-то определенный период (в этом случае говорят о системе индексов с постоянной базой сравнения — ***базисных индексах***);

2) оценивают относительное изменение уровня изучаемого явления по сравнению с предшествующим периодом (получают систему индексов с переменной базой сравнения — ***цепные индексы***).

Рассмотрим системы цепных и базисных индексов цен, физического объема и стоимости продукции (табл. 9.5). Допустим, имеются данные об объеме продукции в натуральном выражении и уровне цен на эту продукцию за пять периодов (номера периодов обозначим 0, 1, 2, 3, 4).

Для индивидуальных индексов цен, индексов физического объема и индексов стоимости продукции справедливо следующее правило:

1) произведение промежуточных по периодам цепных индексов дает базисный индекс последнего периода:

$$i_{4/0} = i_{1/0} \cdot i_{2/1} \cdot i_{3/2} \cdot i_{4/3};$$

2) отношение базисного индекса отчетного периода к базисному индексу предшествующего периода дает цепной индекс отчетного периода:

$$i_{4/3} = i_{4/0} : i_{3/0}.$$

Это правило позволяет применить так называемый цепной метод, т.е. находить неизвестный ряд базисных индексов по известным цепным, и наоборот.

Рассмотрим возможность применения цепного метода исчисления для агрегатных индексов. Имея два базисных агрегатных индекса физического объема с постоянным соизмерителем (цены базисного периода p^0), можно получить цепной индекс отчетного периода:

$$I_{4/3} = \frac{\sum p^0 q^4}{\sum p^0 q^3} = \frac{\sum p^0 q^4}{\sum p^0 q^0} : \frac{\sum p^0 q^3}{\sum p^0 q^0}.$$

Базисный индекс отчетного периода может быть получен путем перемножения соответствующих цепных индексов, если соизмеритель (цена) принимается на уровне одного и того же периода (например, p^0):

$$I_{3/0} = \frac{\sum q^3 p^0}{\sum q^0 p^0} = \frac{\sum p^0 q^1}{\sum p^0 q^0} \cdot \frac{\sum p^0 q^2}{\sum p^0 q^1} \cdot \frac{\sum p^0 q^3}{\sum p^0 q^2}.$$

Использование цепного метода для агрегатных индексов физического объема оказалось возможным благодаря применению в качестве соизмерителя фиксированных (сопоставимых) цен.

Однако, если воспользоваться другим вариантом построения индекса физического объема (см. примечание к табл. 9.5), нельзя получить точного значения базисного индекса путем перемножения соответствующих цепных индексов, так как при исчислении каждого следующего индекса цены меняются: получается ряд индексов с меняющимся соизмерителем:

$$I_{3/0} = \frac{\sum p^0 q^3}{\sum p^0 q^0} \neq \frac{\sum p^0 q^1}{\sum p^0 q^0} \cdot \frac{\sum p^1 q^2}{\sum p^1 q^1} \cdot \frac{\sum p^2 q^3}{\sum p^2 q^2}.$$

Таким образом, при использовании переменных соизмерителей цепной метод для расчета базисных агрегатных индексов физического объема применить нельзя.

Агрегатные индексы качественных показателей, рассчитанные по формуле Пааше, всегда являются индексами с меняющимися весами, так как количество продукции каждый раз принимается на уровне отчетного периода (см. индексы цен с весами отчетного периода в табл. 9.5). Поэтому цепной метод расчета индексов неприменим к таким агрегатным индексам качественных показателей.

Если же воспользоваться формулой Ласпейреса для расчета агрегатных индексов цен при условии постоянных весов q^0 для всех периодов, то базисные индексы могут быть определены на основе цепных, а цепные индексы могут быть рассчитаны с помощью базисных индексов (см. табл. 9.5 — индексы цен с весами базисного периода q^0).

Таблица 9.5

Индексы			Базисные				Цепные			
Физического объема продукции	индивидуальные		$\frac{q^1}{q^0}$	$\frac{q^2}{q^0}$	$\frac{q^3}{q^0}$	$\frac{q^4}{q^0}$	$\frac{q^1}{q^0}$	$\frac{q^2}{q^1}$	$\frac{q^3}{q^2}$	$\frac{q^4}{q^3}$
	агрегатные	в ценах базисного периода	$\frac{\sum p^0 q^1}{\sum p^0 q^0}$	$\frac{\sum p^0 q^2}{\sum p^0 q^0}$	$\frac{\sum p^0 q^3}{\sum p^0 q^0}$	$\frac{\sum p^0 q^4}{\sum p^0 q^0}$	$\frac{\sum p^0 q^1}{\sum p^0 q^0}$	$\frac{\sum p^0 q^2}{\sum p^0 q^1}$	$\frac{\sum p^0 q^3}{\sum p^0 q^2}$	$\frac{\sum p^0 q^4}{\sum p^0 q^3}$
		в ценах отчетного периода	$\frac{\sum p^1 q^1}{\sum p^1 q^0}$	$\frac{\sum p^2 q^2}{\sum p^2 q^0}$	$\frac{\sum p^3 q^3}{\sum p^3 q^0}$	$\frac{\sum p^4 q^4}{\sum p^4 q^0}$	$\frac{\sum p^1 q^1}{\sum p^1 q^0}$	$\frac{\sum p^2 q^2}{\sum p^2 q^1}$	$\frac{\sum p^3 q^3}{\sum p^3 q^2}$	$\frac{\sum p^4 q^4}{\sum p^4 q^3}$
Цен	индивидуальные		$\frac{p^1}{p^0}$	$\frac{p^2}{p^0}$	$\frac{p^3}{p^0}$	$\frac{p^4}{p^0}$	$\frac{p^1}{p^0}$	$\frac{p^2}{p^1}$	$\frac{p^3}{p^2}$	$\frac{p^4}{p^3}$
	агрегатные	с весами базисного периода	$\frac{\sum p^1 q^0}{\sum p^0 q^0}$	$\frac{\sum p^2 q^0}{\sum p^0 q^0}$	$\frac{\sum p^3 q^0}{\sum p^0 q^0}$	$\frac{\sum p^4 q^0}{\sum p^0 q^0}$	$\frac{\sum p^1 q^0}{\sum p^0 q^0}$	$\frac{\sum p^2 q^0}{\sum p^1 q^0}$	$\frac{\sum p^3 q^0}{\sum p^2 q^0}$	$\frac{\sum p^4 q^0}{\sum p^3 q^0}$
		с весами отчетного периода	$\frac{\sum p^1 q^1}{\sum p^0 q^1}$	$\frac{\sum p^2 q^2}{\sum p^0 q^2}$	$\frac{\sum p^3 q^3}{\sum p^0 q^3}$	$\frac{\sum p^4 q^4}{\sum p^0 q^4}$	$\frac{\sum p^1 q^1}{\sum p^0 q^1}$	$\frac{\sum p^2 q^2}{\sum p^1 q^2}$	$\frac{\sum p^3 q^3}{\sum p^2 q^3}$	$\frac{\sum p^4 q^4}{\sum p^3 q^4}$
Стоимости	индивидуальные		$\frac{p^1 q^1}{p^0 q^0}$	$\frac{p^2 q^2}{p^0 q^0}$	$\frac{p^3 q^3}{p^0 q^0}$	$\frac{p^4 q^4}{p^0 q^0}$	$\frac{p^1 q^1}{p^0 q^0}$	$\frac{p^2 q^2}{p^1 q^1}$	$\frac{p^3 q^3}{p^2 q^2}$	$\frac{p^4 q^4}{p^3 q^3}$
	агрегатные		$\frac{\sum p^1 q^1}{\sum p^0 q^0}$	$\frac{\sum p^2 q^2}{\sum p^0 q^0}$	$\frac{\sum p^3 q^3}{\sum p^0 q^0}$	$\frac{\sum p^4 q^4}{\sum p^0 q^0}$	$\frac{\sum p^1 q^1}{\sum p^0 q^0}$	$\frac{\sum p^2 q^2}{\sum p^1 q^1}$	$\frac{\sum p^3 q^3}{\sum p^2 q^2}$	$\frac{\sum p^4 q^4}{\sum p^3 q^3}$

При исчислении цепных агрегатных индексов физического объема (см. табл. 9.5) может быть применена и другая система соизмерителей. Например, при исчислении отдельных индексов используются цены периода, предшествующего отчетному. Тогда получают следующий ряд цепных индексов:

$$I_{1/0} = \frac{\sum q^1 p^0}{\sum q^0 p^0};\ I_{2/0} = \frac{\sum q^2 p^1}{\sum q^1 p^1};\ I_{3/2} = \frac{\sum q^3 p^2}{\sum q^2 p^2};\ I_{4/3} = \frac{\sum q^4 p^3}{\sum q^3 p^3}.$$

То же самое может иметь место и при определении цепных агрегатных индексов цен с весами базисного периода. В качестве веса могут использоваться физические объемы продукции периода, предшествующего отчетному.

Сформулированное выше правило взаимосвязи цепных и базисных индивидуальных индексов в полном объеме применимо к агрегатным индексам стоимости.

9.8. Каковы особенности построения индексных моделей в задачах экономического анализа?

В случае применения индексов в задачах экономического анализа индексы количественных и качественных показателей не рассматриваются изолированно, а исследуются как факторы, определяющие изменение определенного результативного показателя.

Это важное требование увязки индексов в систему объясняет применяемые на практике правила построения агрегатных индексов количественных и качественных показателей.

Допустим, фирма реализует один продукт. Очевидно, что индекс выручки от реализации продукта равен произведению индекса цен и индекса физического объема продукции:

$$i_{pq} = \frac{p_i^1 q_i^1}{p_i^0 q_i^0} = i_{p_i} \cdot i_{q_i}.$$

Такой же должна быть взаимосвязь и между агрегатными индексами: $I_{pq} = I_q \times I_p$,

где $I_{pq} = \dfrac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^0}$.

Какие же агрегатные индексы физического объема и цен будут удовлетворять этому условию? Обратимся к табл. 9.4, где представлены формулы агрегатных индексов Ласпейреса и Пааше. Сопоставив варианты приведенных в таблице индексов, можно видеть, что равенство индекса стоимости произведению агрегатных индексов цен и физического объема продукции соблюдается в двух вариантах сочетания этих индексов:

1) индекса физического объема Ласпейреса и индекса цен Пааше:

$$I_{pq} = \frac{\sum_{i=1}^{n} p_i^0 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^0} \times \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^1} = \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^0};$$

2) индекса физического объема Пааше и индекса цен Ласпейреса:

$$I_{pq} = \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^1 q_i^0} \times \frac{\sum_{i=1}^{n} p_i^1 q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0} = \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^0}.$$

И опять возникает проблема выбора одного из указанных вариантов. Здесь на первый план выдвигается экономическое содержание решаемых с помощью индекса задач. Так, в числителе индекса цен Пааше записана фактическая стоимость продукции отчетного периода, которая сравнивается со стоимостью фактического выпуска по базисным ценам, т.е. в случае роста цен речь идет о реальной дополнительной выручке от реализации произведенной продукций, связанной с увеличением цен производителя. Именно это обстоятельство и обусловливает выбор варианта агрегатного индекса цен Пааше. Тогда, чтобы получить индекс стоимости, агрегатный индекс физического объема должен быть рассчитан по формуле Ласпейреса.

Предпосылкой для проведения анализа в индексной форме является возможность представления результативного экономического показателя с помощью перемножения двух или более определяющих его величину показателей (факторов) или их суммирования. Например, стоимостный объем экспорта может быть представлен в виде произведения уровня внешнеторговых цен p_j и объема экспортных поставок в натуральном выражении q_j[31].

[31] p_j и q_j — соответственно внешнеторговая цена единицы продукции и объем экспорта продукции в натуральном выражении, поставляемого в j-ю страну.

Таким образом, стоимость экспорта зависит от изменения внешнеторговых цен или объема поставок продукции в натуральном выражении либо от одновременного изменения указанных факторов. Поэтому при анализе динамики или выполнения плана по экспорту продукции необходимо показать, в какой мере изменение стоимости экспорта продукции вызвано изменением каждого из этих факторов. С экономической точки зрения небезразлично, какой из этих факторов оказал решающее влияние на увеличение объема экспорта.

Оценивать роль отдельных факторов изменения результативного показателя статистика может путем построения системы взаимосвязанных индексов. Задача состоит в том, чтобы рассчитать изменение сложного показателя при изменении величины только одного фактора так, чтобы величина других факторов была сохранена на определенном постоянном уровне. В основе аналитических индексных расчетов лежит принцип элиминирования изменений величины всех факторов, кроме изучаемого. При построении индексов, оценивающих влияние отдельных факторов на изменение сложного показателя, необходимо иметь в виду, что общий результат изменения этого показателя представляет собой сумму изменения за счет влияния всех исследуемых факторов, формирующих этот показатель.

Сформулируем два дополнительных правила, позволяющих обеспечить выполнение этих условий:

1) при расчете индексов количественных показателей соизмерители принимаются на уровне базисного периода, т.е. расчет ведется по формуле Ласпейреса;

2) при расчете индексов качественных показателей веса в числителе и знаменателе фиксируются на уровне, относящемся к текущему периоду, т.е. используется формула Пааше.

Формулы агрегатных индексов позволяют разделить абсолютный прирост результативного показателя по факторам, например:

$$\Delta pq = \Delta pq_{(q)} + \Delta pq_{(p)}, \qquad (9.15)$$

где Δpq — абсолютный прирост стоимости продукции;

$\Delta pq_{(q)}$ — абсолютный прирост стоимости продукции, обусловленный изменением физического объема продукции;

$\Delta pq_{(p)}$ — абсолютный прирост стоимости продукции, обусловленный изменением уровня цен на продукцию.

Каждая из названных величин абсолютного прироста рассчитывается как разность числителя и знаменателя соответствующего агрегатного индекса:

$$\underbrace{\sum_{i=1}^{n} p_i^1 q_i^1 - \sum_{i=1}^{n} p_i^0 q_i^0}_{\Delta pq} = \underbrace{\left(\sum_{i=1}^{n} p_i^0 q_i^1 - \sum_{i=1}^{n} p_i^0 q_i^0\right)}_{\Delta pq_{(q)}} + \underbrace{\left(\sum_{i=1}^{n} p_i^1 q_i^1 - \sum_{i=1}^{n} p_i^0 q_i^1\right)}_{\Delta pq_{(p)}}.$$

Для иллюстрации воспользуемся данными условного примера о поставках одного вида продукции А на экспорт (табл. 9.6).

Можно ли положительно оценить работу отдела маркетинга данного предприятия? Начнем отвечать на поставленный вопрос с определения динамики общего объема поставок. Количество изделий, поставляемых на экспорт, выросло в отчетном периоде на 100 шт., (1100 − 1000), т.е. на 10%. С этих позиций отделу маркетинга должна быть поставлена положительная оценка.

Далее анализируем динамику внешнеторговых цен со всеми странами—импортерами продукции предприятия (см. графы 5 и 4 табл. 9.6). Индивидуальные индексы цен показывают их рост в отчетном периоде в торговле со всеми странами (см. графу 6 табл. 9.6), что также следует оценить положительно.

Выручка предприятия от поставок продукции на экспорт возросла в отчетном периоде по сравнению с базисным на 15 550 долл., или на 15,755% (см. итоги граф 8 и 7 табл. 9.6).

Таблица 9.6

Динамика экспорта продукции А

Страна-импортер	Объем поставок (в шт.)		Внешнеторговая цена (в долл.)		Индексы динамики цен	Стоимость экспорта (в долл.)		Удельный вес поставок в натуральном выражении	
	в базисном периоде q_j^0	в отчетном периоде q_j^1	в базисном периоде p_j^0	в отчетном периоде p_j^1		в базисном периоде $p_j^0 q_j^0$	в отчетном периоде $p_j^1 q_j^1$	в базисном периоде	в отчетном периоде
1	2	3	4	5	6	7	8	9	10
Болгария	700	450	100	102	1,0200	70 000	45 900	0,7	0,39
Германия	200	250	95	97	1,02105	19 000	24 250	0,2	0,22
Китай	100	450	97	98	1,0103	97 00	44 100	0,1	0,39
Итого	1000	1150				98 700	114 250	1,0	1,00

Примечание. Знак j соответствует индексу страны-импортера.

Однако более детальный анализ динамики стоимости экспорта приводит к несколько иным выводам. По самым высоким ценам осуществляются поставки продукции в Болгарию, тем не менее в отчетном периоде объем поставок (в натуральном выражении) в Болгарию снизился на 35,71% ((450/700)×100% − 100%). Одновременно на 25% возросли поставки в Германию и на 350% в Китай (индексы физического объема продукции равны соответственно 1,25 и 4,50), где внешнеторговые цены ниже. Поскольку речь идет об изменении распределения объема поставок между странами в натуральном выражении, это не могло не отразиться на изменении среднего уровня внешнеторговой цены.

9.9. Что такое индексы переменного состава и влияния структурных сдвигов?

Используя данные о количестве единиц продукта, поставленных в каждую страну (см. графы 2 и 3 табл. 9.6), и уровне внешнеторговых цен (графы 4 и 5 табл. 9.6), рассчитаем среднюю цену поставок продукта А на экспорт в отчетном и базисном периодах по формуле средней арифметической взвешенной:

$$\bar{p}^0 = \frac{\sum_{j=1}^{3} p_j^0 q_j^0}{\sum_{j=1}^{3} q_j^0} = \frac{98\ 700}{1000} = 98{,}7 \text{ долл.};$$

$$\bar{p}^1 = \frac{\sum_{j=1}^{3} p_j^1 q_j^1}{\sum_{j=1}^{3} q_j^1} = \frac{114\ 250}{1150} = 99{,}3478 \text{ долл.}$$

Индекс средних цен экспорта продукта А определяется как соотношение средней цены отчетного и средней цены базисного периодов:

$$I_{\bar{p}} = \frac{\bar{p}^1}{\bar{p}^0} = \frac{\sum_{j=1}^{k} p_j^1 q_j^1}{\sum_{j=1}^{k} q_j^1} : \frac{\sum_{j=1}^{k} p_j^0 q_j^0}{\sum_{j=1}^{k} q_j^0}, \qquad (9.16)$$

где k — число стран — импортеров продукта.

Поскольку величины $\frac{q_j^1}{\sum_{j=1}^{k} q_j^1}$ и $\frac{q_j^0}{\sum_{j=1}^{k} q_j^0}$ отражают распределение экспорта продукта по странам, т.е. географическую структуру экспорта, формула индекса средних цен экспорта может быть записана так:

$$I_{\bar{p}} = \frac{\sum_{j=1}^{k} p_j^1 d_{q_j}^1}{\sum_{j=1}^{k} p_j^0 d_{q_j}^0}. \qquad (9.16а)$$

Величина индекса средних цен экспорта составит 1,00656:

$$I_{\bar{p}} = \frac{99{,}3478}{98{,}7} = 1{,}00656,$$

т.е. прирост средних цен экспорта составил 0,656%. Эта величина значительно ниже, чем прирост цен в торговле с отдельными странами, его величина изменяется от 1,03 до 2,105% (см. графу 6 табл. 9.3). Чем можно объяснить это обстоятельство?

При исчислении средней цены экспорта единицы продукции отчетного периода весами служит количество продукции отчетного периода; при определении средней цены экспорта единицы продукции базисного периода — количество продукции базисного периода. Отношение двух взвешенных средних с меняющимися (переменными) весами, показывающее изменение индексируемой величины, в статистике принято называть ***индексом переменного состава***. Величина этого индекса будет зависеть от изменения уровня внешнеторговых цен при поставках продукции в отдельные страны и от изменения в распределении физического объема экспорта между странами.

Чтобы элиминировать влияние изменений в структуре весов на показатель изменения уровня цен, рассчитывают отношение средних взвешенных цен с одними и теми же весами, т.е. исчисляют индекс цен фиксированного состава. Для этого среднюю внешнеторговую цену единицы продукции в базисном периоде корректируют на структуру фактического объема экспорта продукции. Тогда формула индекса средних цен экспорта фиксированного состава будет такова:

$$I_{\bar{p}} = \frac{\sum_{j=1}^{n} p_j^1 q_j^1}{\sum_{j=1}^{n} q_j^1} : \frac{\sum_{j=1}^{n} p_j^0 q_j^1}{\sum_{j=1}^{n} q_j^1} = \frac{\sum_{j=1}^{n} p_j^1 d_{q_j}^1}{\sum_{j=1}^{n} p_j^0 d_{q_j}^1}. \qquad (9.17)$$

Отношение (9.17) показывает, каково было бы изменение среднего уровня внешнеторговых цен по группе стран, в которые данный вид продукции экспортируется, если бы удельный вес экспорта в разные страны в базисном периоде был таким же, как и в отчетном. Но, как указывалось выше, величина взвешенной средней зависит от двух факторов — изменения отдельных уровней цен (величины p_j по отдельным странам) и от изменения в структуре весов. Поэтому если веса не остаются постоянными, индекс фиксированного состава будет отличаться от индекса переменного состава в меру отношения:

$$I_{\text{перем. состава}} / I_{\text{фикс. состава}} = \left(\frac{\sum_{j=1}^{k} p_j^1 q_j^1}{\sum_{j=1}^{k} q_j^1} : \frac{\sum_{j=1}^{k} p_j^0 q_j^0}{\sum_{j=1}^{k} q_j^0} \right) : \frac{\sum_{j=1}^{k} p_j^1 q_j^1}{\sum_{j=1}^{k} p_j^0 q_j^1}.$$

Указанное отношение принято называть ***индексом влияния структурных сдвигов***. После несложных преобразований получим:

$$I_{\text{влияния структурных сдвигов}} = \frac{\sum_{j=1}^{k} p_j^0 q_j^1}{\sum_{j=1}^{k} q_j^1} : \frac{\sum_{j=1}^{k} p_j^0 q_j^0}{\sum_{j=1}^{k} q_j^0} = \frac{\sum_{j=1}^{k} p_j^0 d_{q_j}^1}{\sum_{j=1}^{k} p_j^0 d_{q_j}^0}. \qquad (9.18)$$

Из приведенной формулы видно, что ***индекс влияния структурных сдвигов*** представляет собой отношение среднего уровня внешнеторговой цены базисного периода, рассчитанного по структуре экспорта определенного вида продукции в отчетном периоде, и фактической средней внешнеторговой цены в базисном периоде. Покажем вычисление индексов цен фиксированного состава и влияния структурных сдвигов на примере (см. данные табл. 9.6).

Агрегатный индекс цен фиксированного состава определяется следующим образом:

$$I_p = \frac{\sum_{j=1}^{k} p_j^1 q_j^1}{\sum_{j=1}^{k} p_j^0 q_j^1} = \frac{\sum_{j=1}^{k} p_j^1 d_{q_j}^1}{\sum_{j=1}^{k} p_j^0 d_{q_j}^0},$$

т.е. расчет индекса может производиться при выборе в качестве веса как объема поставок продукции в натуральном выражении (q_j), так и доли поставок в каждую страну (d_j).

В первом случае величина агрегатного индекса цен рассчитывается следующим образом:

$$I_p = \frac{\sum_{j=1}^{k} p_j^1 q_j^1}{\sum_{j=1}^{k} p_j^0 q_j^1} = \frac{102 \cdot 450 + 97 \cdot 250 + 98 \cdot 450}{100 \cdot 450 + 95 \cdot 250 + 97 \cdot 450} = \frac{114\ 250}{112\ 400} = 1{,}01646.$$

Во втором случае при использовании в качестве веса показателей структуры экспорта в отчетном периоде получим то же значение агрегатного индекса:

$$I_p = \frac{\sum_{j=1}^{k} p_j^1 q_{ij}^1}{\sum_{j=1}^{k} p_j^0 q_{ij}^1} = \frac{102 \cdot 0{,}39 + 97 \cdot 0{,}22 + 98 \cdot 0{,}39}{100 \cdot 0{,}39 + 95 \cdot 0{,}22 + 97 \cdot 0{,}39} = \frac{99{,}34}{97{,}73} = 1{,}01646.$$

Следует обратить внимание на то, что разность числителя и знаменателя двух вариантов индекса цен фиксированного состава будет характеризовать изменение разных показателей. В первом случае эта разность отражает влияние цен на изменение стоимости экспорта, а во втором — влияние этого же фактора на изменение среднего уровня внешнеторговых цен экспорта.

Агрегатный индекс цен фиксированного состава будет равен 101,646%, а это означает, что в среднем по всем странам внешнеторговые цены возросли на 1,646%. Величина этого индекса получилась больше величины индекса переменного состава, что свидетельствует о неблагоприятных структурных сдвигах в экспорте продукции.

Величина индекса влияния структурных сдвигов будет равна 99,02%:

$$I_{\text{влияния структурных сдвигов}} = \frac{100 \cdot 0{,}39 + 95 \cdot 0{,}22 + 97 \cdot 0{,}39}{100 \cdot 0{,}70 + 95 \cdot 0{,}20 + 97 \cdot 0{,}10} =$$

$$= \frac{39 + 20{,}9 + 37{,}83}{70 + 19 + 9{,}7} = \frac{97{,}73}{98{,}70} = 0{,}9902 \text{ (или } 9{,}02\%).$$

Следовательно, за счет изменения в распределении объемов экспорта между странами средняя внешнеторговая цена в отчетном

периоде снизилась на 0,98%, а в абсолютном выражении — на 0,97 долл. Какое влияние оказывает перераспределение физического объема экспорта между странами на величину экспортной выручки? Общая стоимость экспорта $\sum_{j=1}^{k} p_j q_j$ может быть представлена в данном случае величиной $\bar{p} \times \sum_{j=1}^{k} q_j$, где $\bar{p}$ — средняя внешнеторговая цена единицы продукции.

Общее изменение выручки от экспорта продукта будет равно:

$$\Delta pq = \bar{p}^{1} \sum_{j=1}^{k} q_j^1 - \bar{p}^{0} \sum_{j=1}^{k} q_j^0. \qquad (9.19)$$

Абсолютный прирост выручки, обусловленный изменением средних цен экспорта, составит:

$$\Delta pq(\bar{p}) = \left(\bar{p}^{1} - \bar{p}^{0}\right) \sum_{j=1}^{k} q_j^1 = (99{,}3478 - 98{,}7) \cdot 1150 = 745 \text{ долл.}$$

Величина $\sum_{j=1}^{k} q_j$ выступает как вес для качественного показателя — средней цены, а поэтому фиксируется на уровне отчетного периода. Абсолютный прирост выручки, связанный с изменением физического объема поставок, будет равен:

$$\Delta pq\left(\sum_{j=1}^{k} q_j^1\right) = \left(\sum_{j=1}^{k} q_j^1 - \sum_{j=1}^{k} q_j^0\right) \bar{p}^{0} = (1150 - 1000) \cdot 98{,}7 = 14\,805$$

долл.

Поскольку в данном случае оценивается влияние изменения физического объема, соизмеритель фиксируется на уровне базисного периода.

Общий абсолютный прирост экспортной выручки составит 15 550 долл. (745 +14 805). (Сравните с результатом непосредственного расчета по итогам граф 8 и 7 табл. 9.6.)

Учитывая, что средняя внешнеторговая цена зависит от уровня цен экспорта в торговле с каждой страной и от распределения общего физического объема между странами, величину Δpq можно записать так:

$$\Delta pq = \sum_{j=1}^{k} p_j^1 d_j^1 \cdot \sum_{j=1}^{k} q_j^1 - \sum_{j=1}^{k} p_j^0 d_j^0 \cdot \sum_{j=1}^{k} q_j^0. \qquad (9.20)$$

Такая запись выявляет влияние трех факторов на изменение экспортной выручки: изменения внешнеторговых цен (p_j), географической структуры экспорта (d_j) и объема поставок (q_j).

Общее изменение экспортной выручки можно представить в виде суммы трех слагаемых:

$$\Delta pq = \Delta pq(p_j) + \Delta pq(d_j) + \Delta pq(q_j). \qquad (9.20а)$$

В случае трехфакторной модели при оценке влияния одного из факторов на изменения анализируемого показателя два других фактора фиксируются на уровне одного и того же периода. Выбор периода, на уровне которого фиксируются значения факторов (их влияние элиминируется), подчиняется правилу, сформулированному в параграфе 9.8. При этом каждый раз необходимо выяснить, каким показателем (количественным или качественным) по отношению к фактору, влияние которого оценивается, являются все остальные. Так, например, при оценке влияния изменения уровня внешнеторговых цен на изменение экспортной выручки распределение физического объема экспорта по странам $\left(d_{q_j}\right)$ и объем поставок продукта в натуральном выражении (q_j) являются количественными факторами по отношению к уровню внешнеторговой цены, поэтому их величина фиксируется на уровне отчетного периода. Тогда:

$$\Delta pq(\bar{p}) = \left(\sum_{j=1}^{k} p_j^1 d_j^1 - \sum_{j=1}^{k} p_j^0 d_j^1\right) \sum_{j=1}^{k} q_j^1 =$$

$$= (99{,}3478 - 97{,}73) \cdot 1150 = 1860{,}5 \text{ долл.}.$$

Полученная величина показывает, что за счет увеличения внешнеторговых цен при экспорте продукции в отдельные страны, экспортная выручка в отчетном периоде возросла на 1860,5 долл.

При построении ***многофакторных моделей*** целесообразно соблюдать определенную последовательность в записи факторов: в основу должно быть положено экономическое содержание произведений двух смежных факторов. В результате может быть осуществлено преобразование (свертывание) сложной многофакторной модели в модель, содержащую меньшее число факторов, что облегчает обоснование выбора весов или соизмерителей соответствующих показателей. Так, в приводимой модели сумма произведений внешне-

торговой цены на долю поставок в j-ю страну $\sum_{j=1}^{k} p_j q_j$ характеризуют среднюю экспортную цену соответствующего периода. В этой двухфакторной модели p_j является качественным показателем, а d_j — количественным. Следовательно, при определении влияния структуры физического объема товарооборота на изменение экспортной выручки p_j будет фиксироваться на уровне базисного периода. Величина физического объема товарооборота по отношению к средней цене экспорта является количественным фактором, поэтому будет фиксироваться на уровне отчетного периода (фактор d_j, влияние которого оценивается, входит в величину средней цены экспорта).

Таким образом:

$$\Delta pq(d_j) = \left(\sum_{j=1}^{k} p_j^0 d_j^1 - \sum_{j=1}^{k} p_j^0 d_j^0\right)\sum_{j=1}^{k} q_j^1 =$$

$$= (97{,}73 - 98{,}70) \cdot 1150 = -1115{,}5 \text{ долл.}$$

Полученная величина означает, что неблагоприятные изменения в географической структуре физического объема экспорта привели к снижению экспортной выручки на 1115,5 долл.

И наконец, изменение третьего фактора — физического объема экспорта — привело к увеличению экспортной выручки на 14 805 долл.:

$$\Delta pq(d_j) = \sum_{j=1}^{k} p_j^0 d_j^0\left(\sum_{j=1}^{k} q_j^1 - \sum_{j=1}^{k} q_j^0\right) =$$

$$= 98{,}70 \cdot (1150 - 1000) = 14\,805 \text{ долл.}$$

Средняя внешнеторговая цена по отношению к физическому объему экспорта является качественным фактором, поэтому фиксируется на уровне базисного периода. Общее изменение стоимости экспорта составит:

$$\Delta pq = 1860{,}5 - 1115{,}5 + 14\,805 = 15\,550 \text{ долл.},$$

т.е. общая величина экспортной выручки под влиянием всех факторов возросла на 15 550 долл. Однако выполненное разложение абсолютного прироста выручки от экспорта показало, что упущенная выгода в связи с изменением географической структуры экспорта составила 1115,5 долл., а основным фактором, повлиявшим на изменение стоимостного объема экспорта, было увеличение его физического объема.

9.10. Как рассчитывается относительный размер влияния факторов?

Рассмотрим по данным табл. 9.7 разложение темпа прироста стоимости экспорта по определяющим его факторам (расчеты выполнены по данным, приведенным в табл. 9.6).

Основываясь на приведенных в табл. 9.7 данных, определим, на сколько процентов возрастает стоимость экспорта за счет увеличения физического объема экспорта и на сколько — за счет роста уровня средних цен экспорта. Для этого преобразуем абсолютный прирост стоимости экспорта под влиянием изменений каждого из факторов, опираясь на двухфакторную модель взаимосвязи названных выше показателей:

$$V = Q \cdot \bar{p}.$$

Тогда:

$$\Delta V = V^1 - V^0 = Q^1\bar{p}^1 - Q^0\bar{p}^0 =$$
$$= Q^0\bar{p}^0\left(I_Q I_p - 1\right) = V^0\left(I_V - 1\right); \qquad (9.21)$$

$$\Delta V_{(Q)} = Q^1\bar{p}^0 - Q^0\bar{p}^0 = \bar{p}^0\left(I_Q Q^0 - Q^0\right) =$$
$$= \bar{p}^0 Q^0\left(I_Q - 1\right) = V^0\left(I_Q - 1\right); \qquad (9.22)$$

$$\Delta V\left(\bar{p}\right) = Q^1\bar{p}^1 - Q^1\bar{p}^0 = Q^1\left(I_Q\bar{p}^0 - \bar{p}^0\right) =$$
$$= Q^0 I_Q\bar{p}^0\left(I_p - 1\right) = V^0 I_Q\left(I_p - 1\right) = V^0\left(I_V - I_Q\right). \qquad (9.23)$$

Полученные выражения используем для расчета темпов прироста стоимости экспорта, вспомнив, что темп прироста получают путем

Таблица 9.7

Динамика показателей экспорта продукции

Показатель	Условное обозначение	Темп роста (в %)	Темп прироста (в %)
Физический объем экспорта	Q	115,000	15,000
Средняя внешнеторговая цена	P	100,656	0,656
Стоимость экспорта	V	115,755	15,755

деления абсолютного прироста на уровень базисного периода (см. главу 8). Темп прироста выручки общий и за счет каждого из указанных факторов определятся так:

$$\frac{\Delta V}{V^0} \cdot 100\% = (I_V - 1) \cdot 100\% = 115{,}755 - 100{,}000 = 15{,}755\%;$$

$$\frac{\Delta V_{(Q)}}{V^0} \cdot 100\% = (I_Q - 1) \cdot 100\% = 115{,}0 - 100{,}0 = 15{,}0\%;$$

$$\frac{\Delta V(p)}{V^0} \cdot 100\% = (I_V - I_Q) \cdot 100\% = 115{,}755 - 115{,}0 = 0{,}755\%.$$

Таким образом, прирост стоимости экспорта за счет роста физического объема экспорта на 15% составил 15%, а за счет роста средних цен экспорта на 0,656% стоимость экспорта возросла на 0,755%.

При однонаправленном изменении факторов, формирующих результативный показатель, может быть рассмотрена структура темпа прироста результативного показателя, т.е. определена доля каждого фактора в общем приросте:

$$\frac{\Delta V(Q)}{\Delta V} \text{ и } \frac{\Delta V(\overline{p})}{\Delta V}.$$

Используя выражения (9.21) — (9.23), определим относительный размер влияния каждого фактора:

$$\frac{\Delta V(Q)}{\Delta V} = \frac{V^0(I_Q - 1)}{V^0(I_V - 1)} = \frac{I_Q - 1}{I_V - 1}; \qquad (9.24)$$

$$\frac{\Delta V(\overline{p})}{\Delta V} = \frac{V^0(I_V - I_Q)}{V^0(I_V - 1)} = \frac{I_V - I_Q}{I_V - 1}. \qquad (9.25)$$

По данным табл. 9.7 определим, что благодаря изменению физического объема экспорта получено 95,21% всего прироста стоимости экспорта $\left[\frac{1{,}15 - 1}{1{,}15755 - 1} \cdot 100\%\right]$. За счет изменения средних цен экспорта получено 4,79% общего прироста стоимости экспорта $\left[\frac{1{,}15755 - 1{,}15}{1{,}15755 - 1{,}00} \cdot 100\%\right]$.

Результаты расчетов свидетельствуют о том, что преобладающее влияние на увеличение выручки от экспорта продукции оказал рост экспорта в натуральном выражении, а на долю роста цен приходится 4,79% общего прироста экспортной выручки.

9.11. Что такое индексы рынка ценных бумаг?

Традиционным направлением использования индексного метода в статистике развитых стран является ***анализ состояния рынка ценных бумаг.*** Индикаторы рынка ценных бумаг рассчитываются по их видам: акциям, облигациям, опционам и др.

По ***степени обобщения*** исследуемой информации можно выделить следующие показатели рынка ценных бумаг:

1) интегральные, характеризующие состояние исследуемого рынка в целом. Например, известный сводный индекс Доу—Джонса «Индекс-65», рассчитывающийся по акциям 30 крупнейших промышленных корпораций, 20 транспортных и 15 коммунальных компаний;

2) частные, которые дополняют интегральный показатель характеристикой отдельных составных частей (элементов) исследуемого рынка. Например, в дополнение к индексу Доу—Джонса могут быть даны характеристики изменения курса акций отдельных промышленных компаний.

По составу изучаемых объектов интегральные индексы могут охватывать как весь мировой рынок акций, так и его географические секторы и рынок акций отдельных государств. Таковы индексы *MSCI* (*Morgan Stanly Capital International*), например, *Europe 13, North America, World Index.*

Секторные интегральные индексы характеризуют состояние внутринационального рынка. Например, индекс Нью-Йоркской фондовой биржи характеризует «движение» акций всех компаний, котируемых на этой бирже.

Субсекторные интегральные индексы являются составной частью секторного индекса и характеризуют динамику акций, например, промышленных, транспортных или финансовых компаний США.

Таким образом, интегральные индексы могут характеризовать различные сегменты глобального рынка ценных бумаг.

Индексы рынка ценных бумаг — это индексы цен акций, обращающихся на рынке, определяющие динамику их изменения на этом рынке. Индексы могут рассчитываться ежедневно, еженедельно, ежемесячно, ежеквартально, по полугодиям, ежегодно. Изме-

нения значений индексов рассматриваются как показатель спроса на рынке.

Индекс рынка ценных бумаг может использоваться для различных сопоставлений:

1) изменение цен определенных акций можно сравнить с индексом определенного сегмента рынка или с индексом всего рынка и прогнозировать будущее движение цен на акции, а также оценивать спрос на данный вид акций;

2) можно сопоставить изменения цен в различных сегментах рынка и делать выводы о том, какой сектор является наиболее прибыльным для инвесторов в данный момент;

3) для сравнения в аналогичных целях цен на акции в разных странах;

4) для сравнения изменения цен акций мелких и крупных компаний.

Биржевые индексы являются ключевыми показателями для прогнозирования общего положения дел на фондовом рынке и в отдельных отраслях. С помощью индексов ставится задача изучить общие тенденции на рынке, которыми будут руководствоваться инвесторы.

На практике используют четыре методических приема для построения интегральных индексов:

1) рассчитывается темп роста (снижения) средней цены акции, определенной по формуле простой средней арифметической;

2) определяется темп роста (снижения) средневзвешенной цены акций (в качестве веса наиболее часто используется количество обращающихся акций);

3) вычисляется средний арифметический темп прироста (снижения) цены акций;

4) рассчитывается средний геометрический темп прироста (снижения) цены акций.

Рассмотрим эти приемы на условном примере, показывающем расчет индексов цен акций трех компаний. Курс акций, а также количество выпущенных акций приведены в табл. 9.8.

Если воспользоваться первым методом расчета индекса цен на акции, то получим следующий результат:

средняя арифметическая (простая) цена акций в базисном периоде:

$$\overline{p}^0 = \frac{20 + 40 + 100}{3} = 53{,}3 \text{ долл.};$$

в отчетном периоде:

$$\overline{p}^1 = \frac{22 + 38 + 120}{3} = 60 \text{ долл.}$$

Индекс цен акций будет равен 1,1257:

$$I_p = 60/53{,}3 = 1{,}1257,$$

т.е. прирост цен на акции компаний в среднем составил 12,57%.

При втором варианте расчета индексов должно учитываться количество выпущенных акций (q_i):

$$I_p = \frac{\sum_{i=1}^{3} p_i^1 q_i}{\sum_{i=1}^{3} p_i^0 q_i} = \frac{940}{900} = 1{,}0444,$$

т.е. прирост цен составил 4,44%.

По аналогичной методике рассчитывается индекс *Standard and Poors 500*, включающий акции 400 промышленных, 20 транспортных, 40 коммунальных и 40 финансовых корпораций[32]. Эта же методология построения индексов используется при расчетах по опционам и фьючерсам. Также рассчитываются индекс *Wilshire 5000*, включающий акции всех корпораций, котируемых на Нью-Йоркской, Американской фондовой биржах и во внебиржевом обороте.

При использовании третьего метода рассчитывается средний арифметический темп прироста цен акций (см. графу 5 табл. 9.8):

$$\overline{T}_{\text{прир}} = \frac{0{,}10 + (-0{,}05) + 0{,}20}{3} = 0{,}0833.$$

Это означает, что средний прирост цен акций составил 8,33%.

Таблица 9.8

Наименование компании	Курс акций (в долл.)		Количество выпущенных акций (в млн ед.) q_i	Прирост курса акций (в %)	Рыночная стоимость всех акций (в млн долл.)	
	базисный период p_i^0	текущий период p_i^1			базисный период	текущий период
1	2	3	4	5	6	7
«Центр»	20	22	20	+10	400	440
«Омега»	40	38	10	– 5	400	380
«Плюс»	100	120	1	+20	100	120
Итого					900	940

[32] Акции этих корпораций зарегистрированы в основном на Нью-Йоркской фондовой бирже.

И наконец, с помощью средней геометрической получим следующий результат:

$$\bar{k} = \sqrt[3]{(1+0{,}10)(1-0{,}05)(1+0{,}20)} = 1{,}0784,$$

следовательно, можно сделать вывод о среднем темпе прироста цен акций на 7,84%.

Если отвлечься от числовых различий в величине полученных индексов, то можно видеть, что общую тенденцию расчета курса акций названных компаний в текущем периоде они характеризуют однонаправленно, хотя имеет место довольно значительная вариация в величине индексов. Главное назначение интегральных показателей — выявить основное направление, основную тенденцию (тренд) движения курса акций на рынке для характеристики деловой активности в экономике.

На условном примере мы показали основные принципы и методические подходы к построению индексов рынка ценных бумаг. Если рассматривать эту проблему с формально-математических позиций, то для сравнения за отдельные периоды можно выбрать один из вариантов рассмотренных средних, с помощью которых и осуществлять мониторинг цен акций. Однако на практике приходится учитывать целый ряд дополнительных обстоятельств, вносящих существенные коррективы в расчет индексов. К такого рода обстоятельствам относится изменение списка компаний, акции которых формируют индекс; дробление акций компаний; выбор весовых коэффициентов и др.

Одна из особенностей фондовых индексов заключается в том, что базовая цена (или стоимость) приводится к 100 или 1000. Покажем методику расчета сводного фондового индекса Московской межбанковской валютной биржи (ММВБ). Начальной датой расчета сводного фондового индекса ММВБ является 22 сентября 1997 г. На эту дату индекс был принят за 100.

База расчета индекса представляет собой список акций, составляющих фондовый индекс, с указанием количества акций каждого вида. На дату начала расчета индекса база расчета включала акции пяти эмитентов: РАО «ЕЭС», «Ростелеком», Мосэнерго, «Норильский никель» и «ЛУКойл».

Для расчета сводного фондового индекса ММВБ используется следующая формула:

$$I = \frac{1}{D} = \frac{\sum_{i=1}^{N} p_i^t Q_i^t}{\sum_{j=1}^{M} p_j^0 Q_j^0},$$

где N — количество видов акций в базе расчета индекса на день его определения;

Q_i^t — количество акций i-го вида в обращении, определяемое как объем эмиссии акций данного вида для данного эмитента за вычетом пакета акций, находящегося в государственной собственности или контролируемого государством;

p_i^t — средняя взвешенная цена акций i-го вида.

Расчет средней взвешенной цены акций производится следующим образом:

$$p_i = \frac{\sum_{k=2}^{n} p_{ik} Q_{ik}}{\sum_{k=1}^{n} Q_{ik}},$$

где p_{ik} и Q_{ik} — соответственно цена и объем k-й сделки по i-й акции;

n — число сделок в торговой системе с i-й акцией за расчетный период времени;

M — количество видов акций в базе расчета индекса на день определения первого значения (для данного индекса на 22.09.97 г.);

Q_j^0 и p_j^0 — соответственно количество и средневзвешенная цена акций j-го вида на 22.09.97 г.;

D — поправочный коэффициент, корректирующий значение индекса при изменении базы расчета. На начало расчета сводного фондового индекса ММВБ его величина была равна 0,01.

Поправочный коэффициент D сохраняет постоянное значение до тех пор, пока база расчета остается неизменной.

При изменении базы расчета индекса (включение новой акции в реестр или исключение акций из реестра и т.д.) коэффициент D корректируется, чтобы избежать скачка индекса, вызванного изменением базы расчета.

9.12. Что такое индекс потребительских цен и как измеряется инфляция?

В условиях перехода к рыночной экономике большое практическое значение приобретает ***анализ динамики цен на товары и услуги.***

Это обусловлено тем, что показатели, характеризующие инфляционные процессы на рынке потребительских товаров, применяются при решении многих актуальных экономических задач. Для оценки динамики цен на товары используется ***индекс потребительских цен*** (ИПЦ), который иногда называют индексом стоимости жизни. Индекс потребительских цен характеризует изменение во времени общего уровня цен на товары и услуги, приобретаемые населением для непроизводственного потребления. ИПЦ ориентирован на решение следующих основных задач:

- оценку инфляции;
- индексацию доходов.

Для расчета индекса потребительских цен используют данные двух источников информации. Первым источником является наблюдение за изменением цен и тарифов на потребительском рынке, которое проводится с 1992 г. специальной государственной службой при Госкомстате РФ. Эта служба занимается сбором и обработкой данных о розничной торговле, объеме производства, обследованием покупок с использованием данных статистики производства, статистики налогов и торговой статистики. Вторым важным источником информации являются выборочные обследования домашних хозяйств, на основе которых получают данные о структуре фактических потребительских расходов населения. На основании этих источников рассчитывается ИПЦ по фиксированному набору основных потребительских товаров и услуг. Методология расчета индекса является единой для многих стран, что позволяет производить международные сопоставления ИПЦ.

Потребительский набор, на основании которого рассчитывается ИПЦ, состоит из трех групп: продовольственные товары, непродовольственные товары и платные услуги, оказываемые населению. Каждая группа представлена конкретными товарами, услугами или малыми товарными подгруппами. Отбор товаров-представителей производится с учетом их относительной важности для населения, устойчивого наличия в продаже, представительности, с точки зрения отражения динамики цен на однородные товары.

С 1996 г. потребительская корзина в России охватывает 382 позиции товаров-представителей и платных услуг, в том числе по продовольственным товарам — 100 позиций, непродовольственным товарам — 201 и платным услугам — 81 позицию. Наблюдение за ценами по этим товарам и услугам проводится во всех столицах республик, входящих в состав Российской Федерации, краев, областей, округов и выборочно — в районных центрах (15% районных центров каждой области).

ИПЦ определяется с месячной, квартальной периодичностью. Расчет за квартал, период с начала года производится путем перемножения месячных, квартальных ИПЦ.

Расчет ИПЦ (индекса стоимости жизни) выполняется в несколько этапов. Первоначально определяются индивидуальные индексы по отдельным товарам (услугам) для данного города. На основе индивидуальных индексов цен и территориальных весов определяются агрегатные индексы цен отдельных товаров, товарных групп и услуг в целом по региону, экономическому району, Российской Федерации. В качестве территориального веса используется численность населения этой территории или удельный вес численности населения обследуемой территории в общей численности населения России:

$$I_p = \frac{\sum_{j=1}^{k} i_{p_{ij}} N_j}{\sum_{j=1}^{k} N_j} = \sum_{j=1}^{k} i_{p_{ij}} d_{N_j}, \qquad (9.26)$$

где $i_{p_{ij}}$ — индивидуальный индекс цен i-го товара, зарегистрированный на j-й территории;

N_j — численность населения j-й территории;

k — число территориальных единиц.

Используя агрегатные индексы по товарам и услугам по региону, Российской Федерации и привлекая данные об удельном весе расходов на их приобретение в потребительских расходах населения, определяют сводные индексы цен по группам продовольственных, непродовольственных товаров и услуг, а затем и сводный ИПЦ по региону и Российской Федерации в целом.

Определение расчетной формулы сводного ИПЦ связано с задачей выбора между весами текущего и базисного периодов. Удобство практического расчета индекса цен по формуле Ласпейреса сделало ее основной формулой для расчета ИПЦ (см. формулу (9.10)), когда вычисляется отношение стоимости потребительской корзины базисного периода в текущих ценах к ее стоимости в базисном периоде. Следовательно, изменение потребительских цен представлено в виде переоценки физических объемов материальных благ и услуг, входящих в потребительскую корзину.

Практическая реализация методики расчета ИПЦ основана на использовании средней арифметической взвешенной по данным об изменении цен по отдельному товару и о размере потребительских расходов, а не физических объемов потребления. Следовательно, формула (9.10) может быть представлена следующим образом:

$$I_p = \frac{\sum_{i=1}^{n} \frac{p_i^t}{p_i^0} \cdot p_i^0 q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0} = \frac{\sum_{i=1}^{n} i_p \cdot Q_i^0}{\sum_{i=1}^{n} Q_i^0}, \qquad (9.27)$$

где p_i^t — цена i-го товара за t-й период времени;

$Q_i^0 = p_i^0 \times q_i^0$ — стоимость потребительских расходов базисного периода по i-му товару.

Формула (9.27) может быть преобразована следующим образом:

$$I_p = \sum_{i=1}^{n} i_p \cdot d_i^0, \qquad (9.27а)$$

где $d_i^0 = \frac{Q_i^0}{\sum_{i=1}^{n} Q_i^0}$, т.е. представляет собой долю расходов по конкретному i-му товару в общем объеме потребительских расходов базисного периода.

Формула Ласпейреса для расчета ИПЦ используется в большинстве стран. Преимуществом этой формулы является постоянная структура потребительских расходов, а это означает, что для расчета индекса не требуется постоянный пересмотр системы весов. Но на практике при сравнении цен по каждому товару за длительные периоды времени нельзя не учитывать такие факторы, как изменение стереотипов потребления, изменение в предложении товаров и услуг и изменение доходов.

Например, в США ИПЦ (*consumer price index — CPI*) рассчитывается по следующей формуле:

$$I_p = \frac{\sum_{i=1}^{n} p_i^j q_i^j}{\sum_{i=1}^{n} p_i^0 q_i^t}, \qquad (9.28)$$

где q_i^t — физический объем потребления i-го товара, фиксированный на каком-либо конкретном уровне t, или средний уровень потребления за несколько периодов времени.

Использование этой формулы позволяет анализировать изменения в текущем периоде по сравнению с любым другим периодом, а не только с тем годом, когда проводилось обследование потребительских цен.

В России в условиях высокой инфляции структура потребления товаров и услуг используется как постоянная в течение года и пересчитывается каждый год. ИПЦ определяется за каждый месяц и нарастающим итогом с начала года по модифицированной (базисно-взвешенной) формуле Ласпейреса:

$$I_p = \frac{\sum_{i=1}^{n} \frac{p_i^j}{p_i^{j-1}} \cdot p_i^{j-1} \cdot q_i^0}{\sum_{i=1}^{n} p_i^0 \cdot q_i^0}, \qquad (9.29)$$

где $p_i^{j-1} \cdot q_i^0 = q_i^0 \cdot p_i^0 \cdot \frac{p_i^1}{p_i^0} \cdot \frac{p_i^2}{p_i^0} \cdot \ldots \cdot \frac{p_i^{j-1}}{p_i^{j-2}}$.

Например, для третьего месяца:

$$I_{p(2/0)} = \frac{\sum \frac{p_i^2}{p_i^1} \cdot \frac{p_i^1}{p_i^0} \cdot p_i^0 \cdot q_i^0}{\sum_{i=1}^{n} p_i^0 \cdot q_i^0}.$$

ИПЦ является одним из важнейших показателей, характеризующих уровень инфляции, и используется в целях осуществления государственной финансовой политики, пересмотра социальных гарантий.

Уровень инфляции, по существу, представляет собой темп прироста цен. Если в 1999 г. ИПЦ составил 136,5%, то годовой уровень инфляции был равен 36,5%.

А чему был равен средний месячный уровень инфляции в 1999 г.? Для его определения воспользуемся формулой средней геометрической, определив среднемесячный темп роста цен в 1999 г.:

$$\bar{i}_{мес} = \sqrt[12]{1{,}365} = 1{,}0263, \text{ или } 102{,}63\%.$$

Таким образом, средний месячный уровень инфляции в 1999 г. составлял 2,63%.

Для макроэкономического анализа процессов в экономике используется индекс-дефлятор, который оценивает степень инфляции по всей совокупности товаров и услуг. Поскольку расчет индекса-дефлятора должен быть максимально приближен к совокупности товаров, произведенных в отчетном периоде, в большинстве стран мира его определяют по формуле индекса Пааше.

В этой главе мы показали основные формы расчета индексов и направления их использования. Чрезвычайно разнообразны индексы, используемые в анализе различных экономических показателей, их приложение к решению конкретных экономических задач рассматривается в курсах статистики отраслей и макроэкономической статистики.

Приложение 1

Таблица значений функции $f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$

t	0	1	2	3	4	5	6	7	8	9
0,0	3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3725	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	8661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0855	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0203	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
4,0	0001	0001	0001	0000	0000	0000	0000	0000	0000	0000

Приложение 2

Таблица значений функции Лапласа при разных значениях t

$$\Phi(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-\frac{t^2}{2}} dt$$

t	$\Phi(t)$	t	$\Phi(t)$	t	$\Phi(t)$	t	$\Phi(t)$
0,00	0,00000	0,30	0,23582	0,60	0,45149	0,90	0,63188
01	00798	31	24344	61	45814	91	63718
02	01596	32	25103	62	46474	92	64243
03	02393	33	25860	63	47131	93	64763
04	03191	34	26614	64	47783	94	65278
05	03988	35	27366	65	48431	95	65789
06	04784	36	28115	66	49075	96	66294
07	05581	37	28862	67	49714	97	66795
08	06376	38	29605	68	50350	98	67291
09	07171	39	30346	69	50981	99	67783
0,10	0,07966	0,40	0,31084	0,70	0,51607	1,00	0,68269
11	08759	41	31819	71	52230	01	68750
12	09552	42	32552	72	52848	02	69227
13	10348	43	33280	73	53461	03	69699
14	11134	44	34006	74	54070	04	70166
15	11924	45	34729	75	54675	1,05	70628
16	12712	46	35448	76	55275	06	71086
17	13499	47	36164	77	55870	07	71538
18	14285	48	36877	78	56461	08	71986
19	15069	49	37587	79	57047	09	72429
0,20	0,15852	0,50	0,38292	0,80	0,57629	1,10	0,72867
21	16633	51	38995	81	58206	11	73300
22	17413	52	39694	82	58778	12	73729
23	18191	53	40389	83	59346	13	74152
24	18967	54	41080	84	59909	14	74571
25	19741	55	41768	85	60468	15	74986
26	20514	56	42452	86	61021	16	75395
27	21284	57	43132	87	61570	17	75800
28	22052	58	43809	88	62114	18	76200
29	22818	59	44481	89	62953	19	76595

t	$\Phi(t)$	t	$\Phi(t)$	t	$\Phi(t)$	t	$\Phi(t)$
1,20	0,76986	1,55	0,87886	1,90	0,94257	2,25	0,97555
21	77372	56	88124	91	94387	26	97618
22	77754	57	88358	92	94514	27	97679
23	78130	58	88589	93	94639	28	97739
24	78502	59	88817	94	94792	29	97798
25	78870	1,60	0,89040	95	94882	2,30	0,97855
26	79233	61	89260	96	95000	31	97911
27	79592	62	89477	97	95116	32	97966
28	79945	63	89690	98	95230	33	98019
29	80295	64	89899	99	95341	34	98072
1,30	0,80640	65	90106	2,00	0,95450	35	98123
31	80980	66	90309	01	95557	36	98172
32	81316	67	90508	02	95662	37	98221
33	81948	68	90704	03	95764	38	98269
34	81975	69	90897	04	95865	39	98315
35	82298	1,70	0,91087	05	95964	2,40	0,98360
36	82617	71	91273	06	96060	41	98405
37	82931	72	91457	07	96155	42	98448
38	83241	73	91637	08	96247	43	98490
39	83547	74	91814	09	96338	44	98531
1,40	0,83849	75	91988	2,10	0,96427	45	98571
41	84146	76	92159	11	96514	46	98611
42	84439	77	92327	12	96599	47	98649
43	84728	78	92492	13	96683	48	98686
44	85013	79	92655	14	96765	49	98723
45	85294	1,80	0,92814	15	96844	2,50	0,98758
46	85571	81	92970	16	96923	51	98793
47	85844	82	93124	17	96999	52	98826
48	86113	83	93275	18	97074	53	98859
49	86328	84	93423	19	97148	54	98891
1,50	0,86639	85	93569	2,20	0,97219	55	98923
51	86696	86	93711	21	97289	56	98953
52	87140	87	93852	22	97358	57	98983
53	87398	88	93989	23	97425	58	99012
54	87644	89	94124	24	97491	59	99040

Продолжение

t	$\Phi(t)$	t	$\Phi(t)$	t	$\Phi(t)$	t	$\Phi(t)$
2,60	0,99068	2,95	0,99682	3,30	0,99903	3,65	0,99974
61	99095	96	99692	31	99907	66	99975
62	99121	97	99702	32	99910	67	99976
63	99146	98	99712	33	99913	68	99977
64	99171	99	99721	34	99916	69	99978
65	99195	3,00	0,99730	35	99919	3,70	0,99978
66	99219	01	99739	36	99922	71	99979
67	99241	02	99747	37	99925	72	99980
68	99283	03	99755	38	99928	73	99981
69	99285	04	99763	39	99930	74	99982
2,70	0,99307	05	99771	3,40	0,99933	75	99982
71	99327	06	99779	41	99935	76	99983
72	99347	07	99786	42	99937	77	99984
73	99367	08	99793	43	99940	78	99984
74	99386	09	99800	44	99942	79	99985
75	99404	3,10	0,99806	45	99944	3,80	0,99986
76	99422	11	99813	46	99946	81	99986
77	99439	12	99819	47	99948	82	99987
78	99456	13	99825	48	99950	83	99987
79	99473	14	99831	49	99952	84	99988
2,80	0,99489	15	99837	3,50	0,99953	85	99988
81	99505	16	99842	51	99955	86	99989
82	99520	17	99848	52	99957	87	99989
83	99535	18	99853	53	99958	88	99990
84	99549	19	99858	54	99960	89	99990
85	99563	3,20	0,99863	55	99961	3,90	0,99990
86	99576	21	99867	56	99963	91	99991
87	99590	22	99872	57	99964	92	99991
88	99502	23	99876	58	99966	93	99992
89	99615	24	99880	59	99967	94	99992
2,90	0,99627	25	99855	3,60	0,99968	95	99992
91	99639	26	99889	61	99969	96	99992
92	99650	27	99892	62	99971	97	99993
93	99661	28	99896	63	99972	98	99993
94	99672	29	99800	64	99973	99	99993

Таблица случайных чисел

5489	5583	3156	0835	1988	3912	0938	7460	0869	4420
3522	0935	7877	5665	7020	9555	7375	7124	7878	5544
7555	7579	2550	2487	9477	0864	2349	1012	8250	2633
5759	3554	5080	9074	7001	6249	3224	6368	9102	2672
6303	6895	3371	3196	7231	2918	7380	0438	7547	2644
7351	5634	5323	2623	7803	8374	2191	0464	0696	9529
7068	7803	8832	5119	6350	0120	5026	3684	5657	0304
3613	1428	1796	8447	0503	5654	3254	7336	9536	1944
5143	4534	2105	0368	7890	2473	4240	8652	9435	1422
9815	5144	7649	8638	6137	8070	5345	4865	2456	5708
5780	1277	6816	1013	2867	9938	3930	3203	5696	1769
1187	0951	5991	5245	5700	5564	7352	0891	6249	6568
4184	2179	4554	9083	2254	2435	2965	5154	1209	7069
2916	2972	9885	0275	0144	8034	8122	3213	7666	0230
5524	1341	9860	6565	6981	9842	0171	2284	2707	3008
0146	5291	2354	5694	0377	5336	6460	9585	3415	2358
4920	2826	5238	5402	7937	1993	4332	2327	6875	5230
7978	1947	6380	3425	7267	7285	1130	7722	0164	8573
7453	0653	3645	7497	5969	8682	4191	2976	0361	9334
1473	6938	4899	5348	1641	3652	0852	5296	4538	4456
8162	8797	8000	4707	1880	9660	8446	1883	9768	0881
5645	4219	0807	3301	4279	4168	4305	9937	3120	5547
2042	1192	1175	8851	6432	4635	5757	6656	1660	5389
5470	7702	6958	9080	5925	8519	0127	9233	2452	7341
4045	1730	6005	1704	0345	3275	4738	4862	2556	8333
5880	1257	6163	4439	7276	6353	6912	0731	9033	5294
9083	4260	5277	4998	4298	5204	3965	4028	8936	5148
1762	8713	1189	1090	8989	7273	3213	1935	9321	4820
2023	2589	1740	0424	8924	0005	1969	1636	7237	1227
7965	3855	4765	0703	1678	0841	7543	0308	9732	1289
7690	0480	8098	9629	4819	7219	7241	5128	3853	1921
9292	0426	9573	4903	5916	6576	8368	3270	6641	0033
0867	1656	7016	4220	2533	6345	8227	1904	5138	2537
0505	2127	8255	5276	2233	3956	4118	8199	6380	6340
6295	9795	1112	5761	2575	6837	3336	9322	7403	8345
6323	2615	3410	3365	1117	2417	3176	2434	5240	5455
8672	8536	2966	5773	5412	8114	0930	4697	6919	4569
1422	5507	7596	0670	3013	1351	3886	3268	9469	2584
2653	1472	5113	5735	1469	9545	9331	5303	9914	6394
0438	4376	3328	8649	8327	0110	4549	7955	5275	2890
2851	2157	0047	7085	1129	0460	6821	8323	2572	8962
7962	2753	3077	8718	7418	8004	1425	3706	8822	1494
3837	4098	0220	1217	4732	0150	1637	1097	1040	7372
8542	4126	9274	2251	0607	4301	8730	7690	6235	3477
0139	0765	8039	9484	2577	7859	1976	0623	1418	6685
6687	1943	4307	0579	8171	8224	8641	7034	3595	3875
6242	5582	5872	3197	4919	2792	5991	4058	9769	1918
6859	9606	0522	4993	0345	8958	1289	8825	6941	7685
6590	1932	6043	3623	1973	4112	1795	8465	2110	8045
3482	0478	0221	6738	7323	5643	4767	0106	2272	9862

Приложение 4

Значения α-процентных пределов $t_{\alpha,k}$ в зависимости от k степеней свободы и заданного уровня значимости α для распределения Стьюдента

k \ α	10,0	5,0	2,5	2,0	1,0	0,5	0,3	0,2	0,1
1	6,314	12,706	25,452	31,821	63,657	127,3	212,2	318,3	636,6
2	2,920	4,303	6,205	6,965	9,925	14,089	18,216	22,327	31,600
3	2,353	3,182	4,177	4,541	5,841	7,453	8,891	10,214	12,922
4	2,132	2,776	3,495	3,747	4,604	5,597	6,435	7,173	8,610
5	2,015	2,571	3,163	3,365	4,032	4,773	5,376	5,893	6,869
6	1,943	2,447	2,969	3,143	3,707	4,317	4,800	5,208	5,959
7	1,895	2,365	2,841	2,998	3,499	4,029	4,442	4,785	5,408
8	1,860	2,306	2,752	2,696	3,355	3,833	4,199	4,501	5,041
9	1,833	2,262	2,685	2,821	3,250	3,690	4,024	4,297	4,781
10	1,812	2,228	2,634	2,764	3,169	3,581	3,892	4,144	4,587
12	1,782	2,179	2,560	2,681	3,055	3,428	3,706	3,930	4,318
14	1,761	2,145	2,510	2,624	2,977	3,326	3,583	3,787	4,140
16	1,746	2,120	2,473	2,583	2,921	3,252	3,494	3,686	4,015
18	1,734	2,101	2,445	2,552	2,878	3,193	8,428	3,610	3,922
20	1,725	2,086	2,423	2,528	2,845	3,153	3,376	3,552	3,849
22	1,717	2,074	2,405	2,508	2,819	3,119	3,335	3,505	3,792
24	1,711	2,064	2,391	2,492	2,797	3,092	3,302	3,467	3,745
26	1,706	2,056	2,379	2,479	2,779	3,067	3,274	3,435	3,704
28	1,701	2,048	2,369	2,467	2,763	3,047	3,250	3,408	3,674
30	1,697	2,042	2,360	2,457	2,750	3,030	3,230	3,386	3,646
∞	1,645	1,960	2,241	2,326	2,576	2,807	2,968	3,090	3,291

Приложение 5

Значения верхнего α% предела $\chi^2\alpha$ в зависимости от вероятности $P(\chi^2 > \chi^2\alpha)$ и числа степеней свободы χ^2-распределения

Число степеней свободы k	Вероятность $P(\chi^2 > \chi^2\alpha)$							
	0,99	0,98	0,95	0,90	0,80	0,70	0,50	0,30
1	1,64	2,7	3,8	5,4	6,6	7,9	9,5	10,83
2	3,22	4,6	6,0	7,8	9,2	11,6	12,3	13,8
3	4,64	6,3	7,8	9,8	11,3	12,8	14,8	16,3
4	6,0	7,8	9,5	11,7	13,3	13,9	16,9	18,5
5	7,3	9,2	11,1	13,4	15,1	16,3	18,9	20,5
6	8,6	10,6	12,6	15,0	16,8	18,6	20,7	22,5
7	9,8	12,0	14,1	16,6	18,5	20,3	22,6	24,3
8	11,0	13,4	15,5	18,2	20,1	21,9	24,3	26,1
9	12,2	14,7	16,9	19,7	21,7	23,6	26,1	27,9
10	13,4	16,0	18,3	21,2	23,2	25,2	27,7	29,6
11	14,6	17,3	19,7	22,6	24,7	26,8	29,4	31,3
12	15,8	18,5	21,0	24,1	26,2	28,3	31,0	32,9
13	17,0	19,8	22,4	25,5	27,7	29,8	32,5	34,5
14	18,2	21,1	23,7	26,9	29,1	31,0	34,0	36,1
15	19,3	22,3	25,0	28,3	30,6	32,5	35,5	37,7
16	20,5	23,5	26,3	29,6	32,0	34,0	37,0	39,2
17	21,6	24,8	27,6	31,0	33,4	35,5	38,5	40,8
18	22,8	26,0	28,9	32,3	34,8	37,0	40,0	42,3
19	23,9	27,2	30,1	33,7	36,2	38,5	41,5	43,8
20	25,0	28,4	31,4	35,0	37,6	40,0	43,0	45,3
21	26,2	29,6	32,7	36,3	38,9	41,5	44,5	46,8
22	27,3	30,8	33,9	37,7	40,3	42,5	46,0	48,3
23	28,4	32,0	35,2	39,0	41,6	44,0	47,5	49,7
24	29,6	33,2	36,4	40,3	43,0	45,5	48,5	51,2
25	30,7	34,4	37,7	41,6	44,3	47,0	50,0	52,6
26	31,8	35,6	38,9	42,9	45,6	48,0	51,5	54,1
27	32,9	36,7	40,1	44,1	47,0	49,5	53,0	55,5
28	34,0	37,9	41,3	45,4	48,3	51,0	54,5	56,9
29	35,1	39,1	42,6	46,7	49,9	52,5	56,0	58,3
30	36,3	40,3	43,8	48,0	50,9	54,0	57,5	59,7

Значения пяти- и однопроцентных
в зависимости от степеней свободы k_1 и k_2
набраны обычным шрифтом,

k_2 — степени свободы для меньшей дисперсии	k_2	k_1 — степени свободы											
		1	2	3	4	5	6	7	8	9	10	11	12
	1	161	200	216	225	230	234	237	239	241	242	243	244
		4 052	**4 999**	**5 403**	**5 625**	**5 764**	**5 889**	**5 928**	**5 981**	**6 022**	**6 056**	**6 082**	**6 106**
	2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41
		98,49	**99,01**	**99,17**	**99,25**	**99,30**	**99,33**	**99,34**	**99,36**	**99,38**	**99,40**	**99,41**	**99,42**
	3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
		34,12	**30,81**	**29,46**	**28,71**	**28,24**	**27,91**	**27,67**	**27,49**	**27,34**	**27,23**	**27,13**	**27,05**
	4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
		21,20	**18,00**	**16,69**	**15,98**	**15,52**	**15,21**	**14,98**	**14,80**	**14,66**	**14,54**	**14,45**	**14,37**
	5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
		16,26	**13,27**	**12,06**	**11,39**	**10,97**	**10,67**	**10,45**	**10,27**	**10,15**	**10,05**	**9,96**	**9,89**
	6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
		13,74	**10,92**	**9,78**	**9,15**	**8,75**	**8,47**	**8,26**	**8,10**	**7,98**	**7,87**	**7,79**	**7,72**
	7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
		12,25	**9,55**	**8,45**	**7,85**	**7,46**	**7,19**	**7,00**	**6,84**	**6,71**	**6,62**	**6,54**	**6,47**
	8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
		11,26	**8,65**	**7,59**	**7,01**	**6,63**	**6,37**	**6,19**	**6,03**	**5,91**	**5,82**	**5,74**	**5,67**
	9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
		10,56	**8,02**	**6,99**	**6,42**	**6,06**	**5,80**	**5,62**	**5,47**	**5,35**	**5,26**	**5,18**	**5,11**
	10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
		10,04	**7,56**	**6,55**	**5,99**	**5,64**	**5,39**	**5,21**	**5,06**	**4,95**	**4,85**	**4,78**	**4,71**
	11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
		9,85	**7,20**	**6,22**	**5,67**	**5,32**	**5,07**	**4,88**	**4,74**	**4,63**	**4,54**	**4,46**	**4,40**
	12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
		9,33	**6,93**	**5,95**	**5,41**	**5,06**	**4,82**	**4,65**	**4,50**	**4,39**	**4,30**	**4,22**	**4,16**
	13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
		9,07	**6,70**	**5,74**	**5,20**	**4,86**	**4,62**	**4,44**	**4,30**	**4,19**	**4,10**	**4,02**	**3,96**
	14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
		8,86	**6,51**	**5,56**	**5,03**	**4,69**	**4,46**	**4,28**	**4,14**	**4,03**	**3,94**	**3,86**	**3,80**
	15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
		8,68	**6,36**	**5,42**	**4,89**	**4,56**	**4,32**	**4,14**	**4,00**	**3,89**	**3,80**	**3,73**	**3,67**
	16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
		8,53	**6,23**	**5,29**	**4,77**	**4,44**	**4,20**	**4,03**	**3,89**	**3,78**	**3,69**	**3,61**	**3,55**
	17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38
		8,40	**6,11**	**5,18**	**4,67**	**4,34**	**4,10**	**3,93**	**3,79**	**3,68**	**3,59**	**3,52**	**3,45**

верхних пределов уклонения величины F
(пятипроцентные пределы F_5, k_1k_2
однопроцентные F_1, k_1k_2 — жирным)

для меньшей дисперсии												
14	16	20	24	30	40	50	75	100	200	500	∞	k_2
245	246	248	249	250	251	252	253	253	254	254	254	1
6 142	**6 169**	**6 208**	**6 234**	**6 258**	**6 286**	**6 302**	**6 323**	**6 334**	**6 352**	**6 361**	**6 366**	
19,42	19,43	19,44	19,45	19,46	19,47	19,47	19,48	19,49	19,49	19,50	19,50	2
99,43	**99,44**	**99,45**	**99,46**	**99,47**	**99,48**	**99,48**	**99,49**	**99,49**	**99,49**	**99,50**	**99,50**	
8,71	8,69	8,66	8,64	8,62	8,60	8,58	8,57	8,56	8,54	8,54	8,53	3
26,92	**26,83**	**26,69**	**26,60**	**26,50**	**26,41**	**26,35**	**26,27**	**26,23**	**26,18**	**26,14**	**26,12**	
5,87	5,84	5,80	5,77	5,74	5,71	5,70	5,68	5,66	5,65	5,64	5,63	4
14,24	**14,15**	**14,02**	**13,93**	**13,83**	**13,74**	**13,69**	**13,61**	**13,57**	**13,52**	**13,48**	**13,46**	
4,64	4,60	4,56	4,53	4,50	4,46	4,44	4,42	4,40	4,38	4,37	4,36	5
9,77	**9,68**	**9,55**	**9,47**	**9,38**	**9,29**	**9,24**	**9,17**	**9,13**	**9,07**	**9,04**	**9,02**	
3,96	3,92	3,87	3,84	3,81	3,77	3,75	3,72	3,71	3,69	3,68	3,67	6
7,60	**7,52**	**7,39**	**7,31**	**7,23**	**7,14**	**7,09**	**7,02**	**6,99**	**6,94**	**6,90**	**6,88**	
3,52	3,49	3,44	3,41	3,38	3,34	3,32	3,29	3,28	3,25	3,24	3,23	7
6,35	**6,27**	**6,15**	**6,07**	**5,98**	**5,90**	**5,85**	**5,78**	**5,75**	**5,70**	**5,67**	**5,65**	
3,23	3,20	3,15	3,12	3,08	3,05	3,03	3,00	2,98	2,96	2,94	2,93	8
5,56	**5,48**	**5,36**	**5,28**	**5,20**	**5,11**	**5,06**	**5,00**	**4,96**	**4,91**	**4,88**	**4,86**	
3,02	2,98	2,93	2,90	2,86	2,82	2,80	2,77	2,76	2,73	2,72	2,71	9
5,00	**4,92**	**4,80**	**4,73**	**4,64**	**4,56**	**4,51**	**4,45**	**4,41**	**4,36**	**4,33**	**4,31**	
2,86	2,82	2,77	2,74	2,70	2,67	2,64	2,61	2,59	2,56	2,55	2,54	10
4,60	**4,52**	**4,41**	**4,33**	**4,25**	**4,17**	**4,12**	**4,05**	**4,01**	**3,96**	**3,93**	**3,91**	
2,74	2,70	2,65	2,61	2,57	2,53	2,50	2,47	2,45	2,42	2,41	2,40	11
4,29	**4,21**	**4,10**	**4,02**	**3,94**	**3,86**	**3,80**	**3,74**	**3,70**	**3,66**	**3,62**	**3,60**	
2,64	2,60	2,54	2,50	2,46	2,42	2,40	2,36	2,35	2,32	2,31	2,30	12
4,05	**3,98**	**3,86**	**3,78**	**3,70**	**3,61**	**3,56**	**3,49**	**3,46**	**3,41**	**3,38**	**3,36**	
2,55	2,51	2,46	2,42	2,38	2,34	2,32	2,28	2,26	2,24	2,22	2,21	13
3,85	**3,78**	**3,67**	**3,59**	**3,51**	**3,42**	**3,37**	**3,30**	**3,27**	**3,21**	**3,18**	**3,16**	
2,48	2,44	2,39	2,35	2,31	2,27	2,24	2,21	2,19	2,16	2,14	2,13	14
3,70	**3,62**	**3,51**	**3,43**	**3,34**	**3,26**	**3,21**	**3,14**	**3,11**	**3,06**	**3,02**	**3,00**	
2,43	2,39	2,33	2,29	2,25	2,21	2,18	2,15	2,12	2,10	2,08	2,07	15
3,56	**3,48**	**3,36**	**3,29**	**3,20**	**3,12**	**3,07**	**3,00**	**2,97**	**2,92**	**2,89**	**2,87**	
2,37	2,33	2,28	2,24	2,20	2,16	2,13	2,09	2,07	2,04	2,02	2,01	16
3,45	**3,37**	**3,25**	**3,18**	**3,10**	**3,01**	**2,96**	**2,89**	**2,86**	**2,80**	**2,77**	**2,75**	
2,33	2,29	2,23	2,19	2,15	2,11	2,08	2,04	2,02	1,99	1,97	1,96	17
3,35	3,27	3,16	3,08	3,00	2,92	2,86	2,79	2,76	2,70	2,67	2,65	

Приложение 7

Соотношение между r и z' для z' значений от 0 до 5[1]

z'	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0100	0,0200	0,0300	0,0400	0,0500	0,0599	0,0699	0,0798	0,0898
0,1	0,0997	0,1096	0,1194	0,1293	0,1391	0,1489	0,1587	0,1684	0,1781	0,1878
0,2	0,1974	0,2070	0,2165	0,2260	0,2355	0,2449	0,2543	0,2636	0,2729	0,2821
0,3	0,2913	0,3004	0,3095	0,3185	0,3275	0,3364	0,3452	0,3540	0,3627	0,3714
0,4	0,3800	0,3885	0,3969	0,4053	0,4136	0,4219	0,4301	0,4382	0,4462	0,4542
0,5	0,4621	0,4700	0,4777	0,4854	0,4930	0,5005	0,5080	0,5154	0,5227	0,5299
0,6	0,5370	0,5441	0,5511	0,5581	0,5649	0,5717	0,5784	0,5850	0,5915	0,5980
0,7	0,6044	0,6107	0,6169	0,6231	0,6291	0,6352	0,6411	0,6469	0,6527	0,6584
0,8	0,6640	0,6696	0,6751	0,6805	0,6858	0,6911	0,6963	0,7014	0,7064	0,7114
0,9	0,7163	0,7211	0,7259	0,7306	0,7352	0,7398	0,7443	0,7487	0,7531	0,7574
1,0	0,7616	0,7658	0,7699	0,7739	0,7779	0,7818	0,7857	0,7895	0,7932	0,7969
1,1	0,8005	0,8041	0,8076	0,8110	0,8144	0,8178	0,8210	0,8243	0,8275	0,8306
1,2	0,8337	0,8367	0,8397	0,8426	0,8455	0,8483	0,8511	0,8538	0,8565	0,8591
1,3	0,8617	0,8643	0,8668	0,8693	0,8717	0,8741	0,8764	0,8787	0,8810	0,8832
1,4	0,8854	0,8875	0,8896	0,8917	0,8937	0,8957	0,8977	0,8996	0,9015	0,9033
1,5	0,9052	0,9069	0,9087	0,9104	0,9121	0,9138	0,9154	0,9170	0,9186	0,9202
1,6	0,9217	0,9232	0,9246	0,9261	0,9275	0,9289	0,9302	0,9316	0,9329	0,9342
1,7	0,9354	0,9367	0,9379	0,9391	0,9402	0,9414	0,9425	0,9436	0,9447	0,9458
1,8	0,9468	0,9478	0,9498	0,9488	0,9508	0,9518	0,9527	0,9536	0,9545	0,9554
1,9	0,9562	0,9571	0,9579	0,9587	0,9595	0,9603	0,9611	0,9619	0,9626	0,9633
2,0	0,9640	0,9647	0,9654	0,9661	0,9668	0,9674	0,9680	0,9687	0,9693	0,9699
2,1	0,9705	0,9710	0,9716	0,9722	0,9727	0,9732	0,9738	0,9743	0,9748	0,9753
2,2	0,9757	0,9762	0,9767	0,9771	0,9776	0,9780	0,9785	0,9789	0,9793	0,9797
2,3	0,9801	0,9805	0,9809	0,9812	0,9816	0,9820	0,9823	0,9827	0,9830	0,9834
2,4	0,9837	0,9840	0,9843	0,9846	0,9849	0,9852	0,9855	0,9858	0,9861	0,9863
2,5	0,9866	0,9869	0,9871	0,9874	0,9876	0,9879	0,9881	0,9884	0,9886	0,9888
2,6	0,9890	0,9892	0,9895	0,9897	0,9899	0,9901	0,9903	0,9905	0,9906	0,9908
2,7	0,9910	0,9912	0,9914	0,9915	0,9917	0,9919	0,9920	0,9922	0,9923	0,9925
2,8	0,9926	0,9928	0,9929	0,9931	0,9932	0,9933	0,9935	0,9936	0,9937	0,9938
2,9	0,9940	0,9941	0,9942	0,9943	0,9944	0,9945	0,9946	0,9947	0,9949	0,9950
3,0	0,9951									
4,0	0,9993									
5,0	0,9999									

[1] Цифры таблицы являются значениями коэффициента корреляции r, соответствующими значениям z', указанным слева и сверху таблицы.

Приложение 8

Значения коэффициента корреляции рангов Спирмэна для двусторонних пределов уровня значимости α

n \ α	0,20	0,10	0,05	0,02	0,01	0,002
4	0,8000	0,8000				
5	0,7000	0,8000	0,9000	0,9000		
6	0,6000	0,7714	0,8286	0,8857	0,9429	
7	0,5357	0,6786	0,7450	0,8571	0,8929	0,9643
8	0,5000	0,6190	0,7143	0,8095	0,8571	0,9286
9	0,4667	0,5833	0,6833	0,7667	0,8167	0,9000
10	0,4424	0,5515	0,6364	0,7333	0,7818	0,8667
11	0,4182	0,5273	0,6091	0,7000	0,7455	0,8364
12	0,3986	0,4965	0,5804	0,6713	0,7273	0,8182
13	0,3791	0,4780	0,5549	0,6429	0,6978	0,7912
14	0,3626	0,4593	0,5341	0,6220	0,6747	0,7670
15	0,3500	0,4429	0,5179	0,6000	0,6536	0,7464
16	0,3382	0,4265	0,5000	0,5824	0,6324	0,7265
17	0,3260	0,4118	0,4853	0,5637	0,6152	0,7083
18	0,3148	0,3994	0,4716	0,5480	0,5975	0,6904
19	0,3070	0,3895	0,4579	0,5333	0,5825	0,6737
20	0,2977	0,3789	0,4451	0,5203	0,5684	0,6586
21	0,2909	0,3688	0,4351	0,5078	0,5545	0,6455
22	0,2829	0,3597	0,4241	0,4963	0,5426	0,6318
23	0,2767	0,3518	0,4150	0,4852	0,5306	0,6186
24	0,2704	0,3435	0,4061	0,4748	0,5200	0,6070
25	0,2646	0,3362	0,3977	0,4654	0,5100	0,5962
26	0,2588	0,3299	0,3894	0,4564	0,5002	0,5856
27	0,2540	0,3236	0,3822	0,4481	0,4915	0,5757
28	0,2490	0,3175	0,3749	0,4401	0,4828	0,5660
29	0,2443	0,3113	0,3685	0,4320	0,4744	0,5567
30	0,2400	0,3059	0,3620	0,4251	0,4665	0,5479

Приложение 9

Общая структура номенклатуры Гармонизированной системы описания и кодирования товаров (ГС)

Код	Наименование разделов	Группа	Подгруппа	Товарная		
				позиция	подпозиция	субпозиция
1	2	3	4	5	6	7
I	Живые животные и продукция животноводства	5	—	44	115	194
II	Продукты растительного происхождения	9	—	79	220	270
III	Жиры и масла животного или растительного происхождения; продукты их расщепления; приготовленные пищевые жиры; воски животного или растительного происхождения	1	—	22	44	53
IV	Продукция пищевой промышленности; алкогольные и безалкогольные напитки и уксус; табак и его заменители	9	—	56	160	181
V	Минеральные продукты	3	—	67	131	151
VI	Продукция химической и связанных с ней отраслей промышленности	11	19	176	505	759
VII	Пластмассы и изделия из них; каучук и резиновые изделия	2	2	43	138	189
VIII	Кожевенное сырье, кожа, пушнина, меховое сырье и изделия из них; шорно-седельные изделия и упряжь; дорожные принадлежности, сумки и подобные им товары; изделия из кишок (за исключением кетгута)	3	—	21	50	74
IX	Древесина и изделия из древесины; древесный уголь; пробка и изделия из нее; изделия из соломы, альфы и прочих материалов для плетения; корзиночные изделия и другие плетеные изделия	3	—	27	57	79

1	2	3	4	5	6	7
X	Бумажная масса из древесины или из других волокнистых растительных материалов; бумажные и картонные отходы и макулатура; бумага, картон и изделия из них	3	—	41	117	149
XI	Текстиль и текстильные изделия	14	3	149	432	809
XII	Обувь, головные уборы, зонты, трости, трости-сиденья, хлысты, кнуты и их части; обработанные перья и изделия из них; искусственные цветы; изделия из волоса	4	—	20	42	55
XII	Изделия из камня, гипса, цемента, асбеста, слюды и из подобных материалов; керамические изделия; стекло и изделия из него	3	2	49	112	138
XIV	Жемчуг природный или культивированный, драгоценные или полудрагоценные камни, драгоценные металлы, неблагородные металлы, покрытые и плакированные драгоценными металлами, и изделия из них; бижутерия; монеты	1	3	18	40	52
XV	Черные и цветные металлы и изделия из них	11	4	157	418	587
XVI	Машины, оборудование и механизмы; электрическое оборудование; их части: звукозаписывающая и воспроизводящая аппаратура; аппаратура для записи и воспроизведения телевизионного изображения и звука; их части и принадлежности	2	—	133	561	762
XVII	Средства наземного, воздушного и водного транспорта и их части и принадлежности	4	—	38	100	132
XVIII	Приборы и аппараты оптические, фотографические, кинематографические, контрольные, прецизионные, медицинские и хирургические; часы; музыкальные инструменты, их части и принадлежности	3	—	56	184	230

Продолжение

1	2	3	4	5	6	7
XIX	Оружие и боеприпасы: их части и принадлежности	1	—	7	15	17
XX	Разные промышленные товары	3	—	32	108	131
XXI	Произведения искусства, предметы коллекционирования и антиквариат	1	—	6	7	7
	ИТОГО	96	33	1241	3558	5010

Приложение 10

Индекс Доу—Джонса

Дата	DJIA	Дата	DJIA
01 июля	10 971,00	13 сентября	11 030,30
02 июля	11 066,00	14 сентября	10 910,30
05 июля	11 139,00	15 сентября	10 801,40
06 июля	11 139,00	16 сентября	10 737,46
07 июля	11 135,00	17 сентября	10 804,00
08 июля	11 187,00	20 сентября	10 823,00
09 июля	11 127,00	21 сентября	10 598,00
12 июля	11 194,00	22 сентября	10 524,00
13 июля	11 201,00	26 сентября	10 279,36
14 июля	11 175,00	27 сентября	10 303,00
15 июля	11 148,00	28 сентября	10 276,00
16 июля	11 186,00	29 сентября	10 213,48
19 июля	11 210,00	30 сентября	10 336,95
20 июля	10 996,00	01 октября	10 273,00
21 июля	11 003,00	02 октября	10 401,23
23 июля	11 037,50	05 октября	10 400,59
26 июля	10 979,00	06 октября	10 588,34
27 июля	10 979,00	07 октября	10 537,05
30 июля	10 655,15	08 октября	10 649,76
02 августа	10 645,96	11 октября	10 648,18
03 августа	10 677,31	12 октября	10 417,06
04 августа	10 675,00	13 октября	10 232,16
05 августа	10 793,82	14 октября	10 286,61
06 августа	10 714,03	15 октября	10 019,71
09 августа	10 707,70	18 октября	10 116,28
10 августа	10 655,15	19 октября	10 204,93
11 августа	10 787,80	20 октября	10 392,36
12 августа	10 789,39	21 октября	10 297,69
13 августа	10 973,65	22 октября	10 470,25
16 августа	11 046,79	25 октября	10 349,93
17 августа	11 117,08	26 октября	10 302,13
18 августа	10 991,38	27 октября	10 394,89
19 августа	10 963,84	28 октября	10 622,53
20 августа	11 100,61	29 октября	10 729,86
23 августа	11 299,76	01 ноября	10 648,51
24 августа	11 283,30	02 ноября	10 581,84
25 августа	11 326,04	03 ноября	10 609,06
26 августа	11 198,45	04 ноября	10 639,06
27 августа	11 090,17	05 ноября	10 704,48
30 августа	10 831,17	08 ноября	10 718,85
31 августа	10 829,00	09 ноября	10 617,32
01 сентября	10 938,00	10 ноября	10 597,74
02 сентября	10 843,00	11 ноября	10 595,30
03 сентября	10 843,10	12 ноября	10 769,01
06 сентября	11 078,40	15 ноября	10 760,75
07 сентября	11 034,10	16 ноября	10 932,63
08 сентября	11 086,30	17 ноября	10 883,09
09 сентября	11 079,40	18 ноября	11 035,70
10 сентября	11 028,40		

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

1. Ефимова М.Р., Петрова Е.В., Румянцев В.Н. Общая теория статистики: Учебник. М.: ИНФРА – М., 1998.

2. Елисеева И.И., Юзбашев М.М. Общая теория статистики: Учебник. М.: Финансы и статистика, 1998.

3. Теория статистики / Под ред. Р.А. Шмойловой. М.: Финансы и статистика, 1998.

4. Уотшем Т. Дж., Паррамоу К. Количественные методы в финансах: Учеб. пособие для вузов / Пер. с англ. Под ред. М.Р. Ефимовой. М.: Финансы, ЮНИТИ, 1999.

5. Статистический словарь. М.: Финстатинформ, 1996.

Оглавление

ПРЕДИСЛОВИЕ 3

ГЛАВА 1
ПРЕДМЕТ И ЗАДАЧИ СТАТИСТИКИ 5

1.1. Как развивалась статистика и каковы ее задачи? 5
1.2. На какие вопросы должен отвечать статистический показатель? 9
1.3. Как определить предмет и метод статистики? 10
1.4. Каковы основные этапы становления российской государственной статистики? 13
1.5. В чем состоят функции органов государственной статистики? 17
1.6. Каковы отрасли статистической науки? 19

ГЛАВА 2
СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ 23

2.1. В чем заключаются требования, предъявляемые к статистической информации? 23
2.2. В чем заключаются программно-методологические и организационные вопросы статистического наблюдения? 24
2.3. Каковы организационные формы статистического наблюдения? 28
2.4. Какие виды статистического наблюдения существуют? 32
2.5. В чем заключаются основные направления совершенствования статистического наблюдения в современных условиях? 36
2.6. В чем заключаются основные ошибки статистического наблюдения и каковы способы контроля статистических данных? 38

ГЛАВА 3
ГРУППИРОВКА СТАТИСТИЧЕСКИХ ДАННЫХ И ЕЕ РОЛЬ В АНАЛИЗЕ ИНФОРМАЦИИ 41

3.1. В чем состоит значение метода группировок в анализе статистических данных? 41
3.2. Какие виды группировок следует выделить? 42
3.3. Какие группировки называют классификациями? 46
3.4. Как выполнить группировку

по количественным признакам? 47
3.5. Каковы виды и функции статистических таблиц? 50
3.6. Какое значение имеют графики в статистике и каковы общие правила их построения? 54
3.7. Какие виды графиков получили наибольшее распространение? 57

ГЛАВА 4
АБСОЛЮТНЫЕ, ОТНОСИТЕЛЬНЫЕ И СРЕДНИЕ ВЕЛИЧИНЫ 63
4.1. Какова роль статистических показателей в управлении экономикой и в чем состоят общие принципы их построения? 63
4.2. Какие группы обобщающих статистических показателей выделяют? 65
4.3. Каково значение разных видов относительных величин? 68
4.4. В чем состоит роль и значение средних величин? 74
4.5. Как обосновать выбор вида степенной средней? 76
4.6. Каковы основные свойства средней арифметической? 79
4.7. Как обосновать выбор весов при расчете взвешенной средней арифметической? 80
4.8. Что такое средняя гармоническая и средняя геометрическая? 82
4.9. Что такое структурные средние и как их определять по несгруппированным данным? 86

ГЛАВА 5
СТАТИСТИЧЕСКИЕ РАСПРЕДЕЛЕНИЯ И ИХ ОСНОВНЫЕ ХАРАКТЕРИСТИКИ 88
5.1. Что такое вариация признака в совокупности и в чем значение ее изучения? 88
5.2. В чем состоят различия в построении рядов распределения с дискретным и непрерывным характером вариации признака? 90
5.3. Какие группы обобщающих показателей используют для характеристики особенностей рядов распределения? 93
5.4. Как графически представить ряд распределения? 95
5.5. В чем состоят особенности расчета средней арифметической в вариационных рядах? 97
5.6. Как оценить степень неравномерности распределения признака? 99
5.7. Как определить моду и медиану в интервальном ряду распределения? 101
5.8. Как рассчитать показатели вариации (колеблемости) признака? 104
5.9. В чем состоит значение относительных показателей вариации? 110
5.10. На какие составляющие можно разложить общую дисперсию признака? 112
5.11. Как оценить дисперсию альтернативного признака? 115
5.12. Что характеризуют показатели дифференциации? 116
5.13. Что характеризуют показатели формы распределения? 118
5.14. Каковы особенности кривой нормального распределения? 122
5.15. Что такое вариационный ряд и группировка? 125

ГЛАВА 6
ВЫБОРОЧНОЕ НАБЛЮДЕНИЕ 130
6.1. В чем состоят преимущества выборочного наблюдения в сравнении с другими видами статистических наблюдений? 130
6.2. Каковы теоретические основы выборочного метода? 133
6.3. Как определить стандартную и предельную ошибку простой случайной выборки? 135
6.4. Как оценить генеральную среднюю и генеральную долю по данным выборки? 138
6.5. Как определить необходимую численность выборки? 140
6.6. Каковы формы организации выборочного наблюдения? 147
6.7. В чем состоят преимущества механической выборки и как определяется величина ее ошибки? 152
6.8. Что такое малые выборки? 156
6.9. Каковы общие понятия о статистической проверке гипотез? Как выбирают тип критической области? 160
6.10. Как проверяется гипотеза о принадлежности «выделяющихся» наблюдений исследуемой генеральной совокупности? 164
6.11. Как проверяется гипотеза о величине средней арифметической и доли? 169
6.12. Как оценить существенность расхождения двух выборочных средних или долей? 171

ГЛАВА 7
КОРРЕЛЯЦИОННАЯ СВЯЗЬ И ЕЕ СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ 176
7.1. Что такое корреляционная связь и какой круг вопросов она изучает? 176
7.2. Каковы основные предпосылки применения корреляционно-регрессионного метода анализа? 178
7.3. Какие приемы используются для выявления наличия корреляционной связи между двумя признаками? 180
7.4. Как измерить степень тесноты корреляционной связи в случае парной зависимости? 186
7.5. Как выполнить оценку существенности линейного коэффициента корреляции? 190
7.6. Что такое эмпирическое корреляционное отношение и какова область его применения? 192
7.7. Какие показатели используют для измерения степени тесноты связи между качественными признаками? 196
7.8. Как рассчитываются значение уравнения регрессии и параметры парной регрессии? 205
7.9. Как проводится анализ параметров парной регрессии? 208
7.10. Каковы способы расчета средней квадратической ошибки и ее роль в оценке надежности уравнения регрессии? 212
7.11. Как осуществить прогноз значений результативного признака по уравнению регрессии? 214
7.12. Как оценить роль отдельных факторов в формировании результативного показателя по уравнению множественной зависимости? 218

ГЛАВА 8
РЯДЫ ДИНАМИКИ 227
8.1. Каковы виды рядов динамики? 227
8.2. Каковы показатели изменения уровней временного ряда и методы их исчисления? 230
8.3. Каковы средние характеристики ряда динамики? 235
8.4. Каковы характеристика основной тенденции развития и способы ее выявления? 241

8.5. Какова роль аналитического выравнивания уровней временного ряда? 245
8.6. Как выполнить прогноз по уравнению тренда? 250
8.7. Что такое сезонная неравномерность и какова ее характеристика? 256
8.8. Какова аналитическая модель проявления сезонных колебаний? 263
8.9. Что такое модели авторегрессии? 265

ГЛАВА 9
ИНДЕКСЫ И ИХ ИСПОЛЬЗОВАНИЕ В ЭКОНОМИКО-СТАТИСТИЧЕСКИХ ИССЛЕДОВАНИЯХ 269
9.1. Что такое индексы и каково значение индексного метода анализа? 269
9.2. На каких принципах базируется расчет агрегатных индексов количественных показателей? 271
9.3. Какие виды средних индексов можно использовать для обобщения динамики количественных показателей? 277
9.4. В чем состоит различие агрегатных индексов цен и других качественных показателей, рассчитанных по формулам Ласпейреса и Пааше? 281
9.5. Что такое средние взвешенные индексы качественных показателей? 284
9.6. Каковы принципиальные основы выбора формы индекса? 287
9.7. Какова взаимосвязь цепных и базисных индексов? 289
9.8. Каковы особенности построения индексных моделей в задачах экономического анализа? 292
9.9. Что такое индексы переменного состава и влияния структурных сдвигов? 296
9.10. Как рассчитывается относительный размер влияния факторов? 303
9.11. Что такое индексы рынка ценных бумаг? 305
9.12. Что такое индекс потребительских цен и как измеряется инфляция? 309

Приложения 315
РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА 330

СТАТИСТИКА

Учебное пособие

Редактор *И.В. Башина*
Корректор *Н.В. Андрианова*
Компьютерная верстка *С.М. Майоров*
Оформление серии *Е.А. Дони*

ЛР № 070824 от 21.01.93.

Сдано в набор 10.08.2000. Подписано в печать 20.09.2000
Формат 60x88/16. Бумага типографская № 2. Гарнитура «NewtonC».
Печать офсетная. Усл. печ. л. 20,58. Уч.-изд. л. 19,71.
Доп. тираж 4 000 экз. Заказ № 6504.
Цена свободная.

Издательский Дом «ИНФРА-М»
127214, Москва, Дмитровское ш., 107.
Тел.: (095) 485-71-77
Факс: (095) 485-53-18. Робофакс: (095) 485-54-44
E-mail: books@infra-m.ru
http://www.infra-m.ru

Отпечатано в полном соответствии с качеством
предоставленных диапозитивов в Тульской типографии.
300600, г. Тула, пр. Ленина,109.